科學 专著丛书

非线性时间序列分析

NON-LINEAR TIME SERIES ANALYSIS

安鸿志 陈 敏 著

上海科学技术出版社

科学专著丛书

非线性时间序列分析

安鸿志　陈　敏　著

上海科学技术出版社

责任编辑　赵序明

科学专著丛书
非线性时间序列分析
安鸿志　陈　敏　著
上海科学技术出版社出版、发行
(上海瑞金二路450号)
新华书店上海发行所经销　　常熟市印刷四厂印刷
开本 787×1092　小 1/16　印张 20.75　插页 4　字数 332 000
1998年8月第1版　　1998年8月第1次印刷
印数：1—1 200
ISBN 7-5323-4373-1/O・207
定价：38.00元

内 容 提 要

本书概述了非线性时间序列分析的基本内容以及近代的重要进展。全书分为两部分：第一部分为模型概论，即第1章至第4章，其中包括了：比较线性与非线性时间序列；概述与非线性时间序列有关的非线性动态系统；介绍各种非线性时间序列模型；论述这些模型的平稳解、遍历性、高阶矩和可逆性。第二部分介绍统计方法，即第5章至第9章，包括非线性时间序列预报方法；非线性性检验方法；非线性时间序列建模方法；与其相关领域（如混沌、分形和神经网络）之间的联系；若干实例应用分析。

本书适合于作大学本科高年级学生和研究生的教材或参考书；也是从事时间序列分析研究工作人员及使用时间序列方法分析数据资料工作者的重要参考书。

《科学专著丛书》序

如果说科学研究论文是创造性科学工作的发表性记录，那末科学技术学术专著则是创造性科学工作的总结性记录。前者注重的是优先权，后者注重的是系统化。

在大量科学研究的基础上，对一个专题或一个领域的研究成果，作系统的整理总结，著书立说，乃是科学研究工作不可少的一个组成部分。著书立说，既是丰富人类知识宝库的需要，也是探索未知领域、开拓人类知识新疆界的需要。特别是在科学各门类的那些基本问题上，一部优秀的学术专著常常成为本学科或相关学科取得突破性进展的基石。所以，科学技术学术专著的著述和出版是一项十分重要的工作。

近 20 年来，中国的科学事业有了迅速的发展，涌现了许多优秀的科学研究成果，为出版学术专著提供了坚实的基础。值此 20 世纪 90 年代，在出版学术专著方面，中国的科学界和出版界都在抓紧为本世纪再加些积累，为迎接新世纪多作些开拓。我高兴地看到，作为这种努力的一个部分，《科学》杂志的出版者——上海科学技术出版社推出了这套《科学专著丛书》。

上海科学技术出版社是科学技术界熟悉和信赖的一家出版社，历来注重科学技术学术专著的出版。《科学》杂志的编者组织编辑学术系列丛书，也不是第一次。在本世纪三四十年代，就曾推出过《科学丛书》，其中不乏佳作，对当时的学术研究起了很好的作用。

《科学》在中国是一份历史最长的综合性科学刊物，80 年来与科学技术界建立了广泛的密切联系。现在推出的这套《科学专著丛书》正是这种联系的产物。我相信，加强这种联系，著者与编者、出版者，科技界与出版界共同努力，精心选题，精心编辑，精心出版，一定能使这套专著丛书反映

出中国科学技术研究的最新水平，为本世纪多留下几本中国学者的优秀专著，为迈向新世纪多铺下几块引路的基石！

周光召

（《科学》杂志编委会主编）

1994年8月

前　言

在过去的半个多世纪里，时间序列分析得到了迅速发展。特别是对线性时间序列分析的研究，已经取得了系统和丰富的成果。但是对于非线性时间序列分析的研究，仅在近二十多年里才逐渐被重视起来。

线性时间序列仅是时间序列的一种特殊类型，而且是最简单的一类。尽管如此，五十多年来，被研究得最深入，应用也最广泛的，也仅是有限参数型线性时间序列模型，例如熟知的自回归滑动平均模型。所以说，非线性时间序列分析是一个极其广泛的研究范畴，有着极为丰富和深刻的研究内容。

非线性时间序列，哪怕是很简单的非线性时间序列，它们能显示出的奇异现象，足以令人新奇不已。况且，非线性时间序列分析与许多非线性科学领域有着密切的联系，例如与混沌、分形和神经网络等领域均有密切的关系。这说明了非线性时间序列能表现出更丰富、更复杂的客观现象，比线性时间序列有着更广阔的应用背景。实际上，在很多领域，例如在经济、环境、生物和气象等领域，已经显示出非线性时间序列分析的应用背景。因此，近年来对非线性时间序列分析的研究受到广泛关注。

研究非线性时间序列，很难仿效或沿用已有的线性时间序列的理论与方法，而且所面临的困难也远远超过线性情况。正像在其他非线性科学领域一样，人们遇到了许多新的激动人心的挑战。

近年来，关于非线性时间序列分析的文献，正在不断丰富。本书作者在了解有关文献的同时，也从事了某些理论和应用研究。并且在非线性时间序列模型的平稳解、遍历性等理论方面，以及非线性性检验方法和随机条件方差模型的应用方面，取得了某些研究成果。这些研究工作使作者对非线性时间序列分析的丰富内涵，有了较新的和深入的了解。这促使作者写出此书，以期吸引更多的同仁学者关心非线性时间序列分析的应用与发展。

在此，我们要感谢汤家豪教授的鼓励和帮助，提供许多最新的论著的

预印本，使我们较及时地了解非线性时间序列研究的最新进展。作者还要感谢香港中文大学统计系王永雄教授，受他的邀请，本书第一作者曾在该系研究生的讨论班上讲过本书的多数章节，并在该系访问期间完成本书初稿。作者还要感谢曾与他们共同组织参加非线性时间序列讨论班的同事们，书中许多内容曾与他们交流或讨论过。此外，还应当感谢北京大学概率统计系谢衷洁教授，他细心审阅了书稿，并提出许多宝贵意见。

该书第一作者多年受到国家自然科学基金委的资助，此书是属于天元重点基金资助项目的科研成果之一，一并在此致谢。

由于作者知识所限，书中谬误之处难免，恳望学者同仁不吝批评指正。

作 者

1996 年 3 月于北京中国科学院

应用数学研究所

符 号 说 明

γ_k	宽平稳时间序列的自协方差
$\gamma_{k,j}$	时间序列的三阶矩函数
$\hat{x}_n(l)$	l 步条件期望预报
$\hat{x}_n^*(l)$	l 步投影预报
$\|\cdot\|$	Euclid 范数
$\|\cdot\|_v$	与 Euclid 范数 $\|\cdot\|$ 等价的某一向量范数
$\boldsymbol{T}_t(\boldsymbol{x}_0)$	以 $\boldsymbol{x}_0$ 为初始值用变换 T 迭代 t 步的值
$L(\boldsymbol{x}_0,\boldsymbol{T})$	迭代序列$\{\boldsymbol{T}_t(\boldsymbol{x}_0),t\geqslant 1\}$的所有可能的极限值全体
$L(\boldsymbol{T})$	变换 T 的吸引集
$r_t(\boldsymbol{x}_0,B)$	点 $T_t(\boldsymbol{x}_0)$到集合 B 的普通 Euclid 距离
$A_T(B)$	变换 T 的吸引元 B 的吸引域
$A\geqslant B$	$A-B$ 是半正定 Hermite 矩阵
$\rho(A)$	矩阵 A 的谱半径，即 A 的特征值最大模
A^τ	矩阵 A 的转置
R^m	m 维实数域空间
$\mathscr{B}_m$	R^m 中的 Borel σ-代数
$P(\boldsymbol{x},A)$	$(R^m,\mathscr{B}_m)$上的转移概率
$I(\boldsymbol{x}\in A)$	集合 $A\in R^m$ 的示性函数
$\|F-G\|_\tau$	分布 F 与 G 之差的全变差
$\mathrm{Vec}(A)$	矩阵 A 的拉直运算
$A\otimes B$	矩阵 A 与 B 的 Kronecker 积
$\xrightarrow{p}$	依概率收敛
$\xrightarrow{d}$	依分布收敛
a. s.	概率为 1 地成立

$d_k(A)$	R^m 中有界集合 A 的 Kolmogorov 容度
$\triangleq$	表示"定义"或"记为"
$[x]$	取实数 x 的整数部分
$\mathscr{F}^x_{a,b}$	由$\{x_s, a\leqslant s\leqslant b\}$生成的 σ-代数
$\mathscr{F}^x_t$	由$\{x_s, s\leqslant t\}$生成的 σ-代数
$\varepsilon_t\|\mathscr{F}^x_{t-1}$	给定 $\mathscr{F}^x_{t-1}$的条件下,ε_t 的条件分布
μ_m	R^m 空间的 Lebesgue 测度

目　录

第 1 章

浅释线性与非线性时间序列

本章的目的是简要地介绍线性时间序列与非线性时间序列的联系，彼此相似之处以及两者间的实质性差异. 通过对线性时间序列基本描述方法的简略回顾以及与非线性时间序列的比较，概述研究非线性时间序列所涉及的基本内容，现在所使用的一些描述方法以及所面临的主要困难.

§1.1 平稳时间序列

从较广的意义上说，时间序列是指被观测到的依时间次序排列的数据序列. 从概率论角度来看，用随机过程来描述最为合适. 随机过程被定义为一簇随机变量，即$\{x_t, t \in T\}$，其中 T 表示时间 t 的变化范围. 对每个固定的时刻 t，x_t 是一个一元随机变量，这些随机变量的全体就构成一个随机过程. 当 $T=\{0,\pm 1,\pm 2,\cdots\}$时，随机过程$\{x_t, t \in T\}$可写成$\{x_t, t=0,\pm 1,\pm 2,\cdots\}$，称之为*随机序列*. 由于 t 代表时间，所以此类随机序列也称为*时间序列*. 这种定义对线性和非线性时间序列都是适用的.

由上述可见，时间序列是一类特殊的随机过程——离散时间的随机过程. 因此，时间序列的概率结构是被其有限维分布簇所确定的. 粗略地说，当对每个自然数 m 和每组整数 $t_1, t_2, \cdots, t_m$，$(x_{t_1}, x_{t_2}, \cdots, x_{t_m})$的 m 维联合分布

$$P\{x_{t_1} < z_1, x_{t_2} < z_2, \cdots, x_{t_m} < z_m) = F(z_1, \cdots, z_m; t_1 \cdots, t_m)$$

都被给定时，时间序列$\{x_t\}$的概率结构就被完全确定. 详细的严格定义见文献[125]. 这一定义对线性与非线性时间序列都适用.

平稳时间序列，又简称*平稳序列*，是一类最重要的时间序列. 对这类时间序列，其统计分析方法已有相当丰富的知识，应用也最为广泛. 而且，许多非平稳序列的统计分析方法也是借助于与平稳序列的联系而建立起来的. 例如，平稳序列的求和序列为非平稳序列，对它的研究即可借用平

稳序列的分析方法. 所以,平稳序列在时间序列分析中起着重要作用. 时间序列的平稳性概念,并不与时间序列的线性性直接联系.

定义 1.1.1 时间序列$\{x_t\}$称为*严平稳的*,如果对任何正整数 m 和整数 $t_1<t_2<\cdots<t_m$,此序列中的随机变量 $x_{t_1+s},x_{t_2+s},\cdots,x_{t_m+s}$的联合分布与整数 s 无关,即

$$F(z_1,\cdots,z_m;t_1+s,\cdots,t_m+s)=F(z_1,\cdots,z_m;t_1,\cdots,t_m). \tag{1.1.1}$$

以上定义表明,严平稳序列的任何有限维联合分布对于时间的平移是不变的. 关于时间序列的平稳性还有另一种不同意义的定义.

定义 1.1.2 时间序列$\{x_t\}$称为*宽平稳的*,如果 x_t 具有有穷的二阶矩,而且满足以下条件:

(1) $\mu_t=Ex_t=c$;

(2) $\gamma(t,s)=E(x_t-c)(x_s-c)$
$\quad=\gamma(t-s,0)$.

定义 1.1.2 表明,宽平稳序列的均值与自协方差函数随时间的平移而不变. 另外,其中条件(2)表明,宽平稳序列的自协方差函数 $\gamma(t,s)$只与 $t-s$ 有关,故 $\gamma(s,t)$实际上是被一元函数所确定. 以后,我们用如下记号表示宽平稳序列的自协方差函数

$$\gamma_{t-s}=E(x_t-c)(x_s-c)=\gamma(t-s,0).$$

由上述两个定义可知,一个严平稳序列未必有二阶矩,因而不一定是宽平稳序列. 反之,一个宽平稳序列的分布未必随时间的推移而不变,因此,也不一定是严平稳序列. 为说明两种不同意义的平稳性的联系,先介绍如下的正态时间序列的定义.

定义 1.1.3 如果时间序列$\{x_t\}$的任何有限维分布都是多元正态分布,则称$\{x_t\}$为*正态时间序列*,亦简称*正态序列*.

正态序列与平稳序列是两种不同的概念. 正态序列可以是平稳的,也可以是非平稳的. 而且,无论在严平稳或宽平稳意义下,都是如此. 以下性质说明了严平稳与宽平稳的一般联系.

性质 1.1.4 如果$\{x_t\}$是严平稳序列,且 $Ex_t^2<\infty$,则$\{x_t\}$必是宽平稳序列. 反之,不一定成立. 如果$\{x_t\}$是正态序列,则$\{x_t\}$为严平稳序列与宽平稳序列是等价的.

此性质在时间序列分析中是熟知的结果. 关于宽平稳序列的自协方差函数的性质,有如下熟知的结果.

性质 1.1.5 宽平稳序列$\{x_t\}$的自协方差函数 γ_k 具有以下性质：

(1) $|\gamma_k| \leqslant \gamma_0$；

(2) $\gamma_k = \gamma_{-k}$；

(3) 对任何自然数 m 和整数 $t_1 < t_2 < \cdots < t_m$ 及实数 $c_1, c_2, \cdots, c_m$，有

$$\sum_{i=1}^{m}\sum_{j=1}^{m} c_i c_j \gamma_{t_i - t_j} \geqslant 0.$$

性质 1.1.6 $\{\gamma_k, k = 0, \pm 1, \pm 2, \cdots\}$为某一宽平稳序列$\{x_t\}$的自协方差函数，当且仅当

$$\gamma_k = \int_{-\pi}^{\pi} \mathrm{e}^{ik\lambda} \mathrm{d}F(\lambda), \tag{1.1.2}$$

其中 $F(\lambda)$为一单调非降左连续函数，且总变差是有穷的. 称 $F(\lambda)$ 为$\{x_t\}$的*谱分布函数*，而(1.1.2)式称为*自协方差函数的谱表示*.

定义 1.1.7 设$\{x_t\}$是宽平稳序列，$F(\lambda)$是$\{x_t\}$的谱分布函数，如果存在非负可积函数 $f(\lambda)$，使得

$$F(\lambda) = \int_{-\pi}^{\lambda} f(\lambda) \mathrm{d}\lambda,$$

则称 $f(\lambda)$ 为$\{x_t\}$的*谱密度函数*.

性质 1.1.8 设$\{x_t\}$是宽平稳序列，其自协方差函数 $\gamma(k)$满足

$$\sum_{k=-\infty}^{\infty} |\gamma_k| < \infty, \tag{1.1.3}$$

则$\{x_t\}$必有非负谱密度函数 $f(\lambda)$，且 $f(\lambda)$和$\{\gamma_k\}$之间是 Fourier 变换的关系：

$$\begin{cases} f(\lambda) = \dfrac{\mathrm{d}F(\lambda)}{\mathrm{d}\lambda} = \dfrac{1}{2\pi}\sum\limits_{k=-\infty}^{\infty} \gamma_k \mathrm{e}^{ik\lambda}; \\ \gamma_k = \displaystyle\int_{-\pi}^{\pi} \mathrm{e}^{ik\lambda} f(\lambda) \mathrm{d}\lambda. \end{cases} \tag{1.1.4}$$

以上性质的证明可见文献[20].

由上述性质可知，宽平稳序列的自协方差函数与谱分布函数具有等价的描述能力. 它们都被序列的一维和二维分布 $F(z;t)$和 $F(z_1, z_2; t_1, t_2)$所决定. 反之，除在正态时间序列情况下，从自协方差函数不能唯一地决定序列的高维分布，从而也就不能确定序列的全部概率结构. 所以说，自协方差函数或谱分布函数都不是序列的全部概率结构信息的表现. 确切地说，它们只是序列的一、二阶矩的结构表征.

一个最简单的例子是

$$x_t = \varepsilon_t \varepsilon_{t-1}, \tag{1.1.5}$$

其中$\{\varepsilon_t\}$为独立正态 $N(0,1)$随机序列. 由(1.1.5)式易见,

$$\begin{aligned} &Ex_t = E\varepsilon_t\varepsilon_{t-1} = E\varepsilon_t E\varepsilon_{t-1} = 0, \\ &Ex_t^2 = E\varepsilon_t^2\varepsilon_{t-1}^2 = E\varepsilon_t^2 E\varepsilon_{t-1}^2 = 1, \\ &Ex_t x_{t+k} = E\varepsilon_t\varepsilon_{t-1}\varepsilon_{t+k}\varepsilon_{t+k-1} \\ &\qquad = E\varepsilon_{t+k} E\varepsilon_t\varepsilon_{t-1}\varepsilon_{t+k-1} = 0 \quad (k \geqslant 1). \end{aligned}$$

以上诸式表明,(1.1.5)式的$\{x_t\}$为宽平稳序列. 而依$\{\varepsilon_t\}$的定义,它也满足相同的性质,即

$$E\varepsilon_t = 0,\ E\varepsilon_t^2 = 1,\ E\varepsilon_t\varepsilon_{t+k} = 0 \quad (k \geqslant 1),$$

即$\{x_t\}$和$\{\varepsilon_t\}$具有相同的均值和自协方差函数. 显然,它们的分布结构不相同.

由于本书主要讨论严平稳性,为了简便,我们限定平稳性专指严平稳性. 本节所讨论的时间序列的平稳性概念,无论对线性时间序列而言,还是对非线性时间序列而言,都是适用的,并且具有重要意义. 对线性时间序列来说,宽平稳序列起着重要作用,而对非线性时间序列来说,我们将更注重于平稳序列的讨论.

§ 1.2 线性时间序列

相互独立且具有相同分布的随机变量序列是一类非常特殊的时间序列. 它在时间序列分析中占有非常重要的地位. 对此类序列还常假定为零均值和有穷方差,在本书中我们称之为*白噪声序列*,常用$\{\varepsilon_t\}$表示之. 白噪声序列是一种概率结构十分简单的平稳序列. 许多重要的、有用的时间序列是由此类序列变换而来的. 本节将要讨论的线性序列就与白噪声序列有密切关系.

定义 1.2.1 如果随机序列$\{x_t\}$可表为

$$x_t = \sum_{j=0}^{\infty} \beta_j \varepsilon_{t-j}, \tag{1.2.1}$$

其中系数序列$\{\beta_j\}$满足

$$\sum_{j=0}^{\infty} \beta_j^2 < \infty, \tag{1.2.2}$$

而$\{\varepsilon_t\}$是白噪声序列,满足

$$E\varepsilon_t = 0,\ E\varepsilon_t^2 = \sigma^2 < \infty, \tag{1.2.3}$$

则称$\{x_t\}$为线性序列，又称(1.2.1)式为$\{\varepsilon_t\}$的无穷滑动平均.

关于线性序列的研究，已有很丰富的文献，如文献[16]、[20]、[57]、[125]、[133]、[135]等. 在线性序列中，应用最广泛的是线性自回归滑动平均(Linear Autoregression and Moving Average)序列，其定义如下：

定义 1.2.2 假定时间序列$\{x_t\}$由如下随机差分方程所确定：

$$x_t = \varphi_1 x_{t-1} + \cdots + \varphi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}, \quad (1.2.4)$$

其中$\{\varepsilon_t\}$为白噪声序列，满足(1.2.3)式，且ε_t与$\{x_s, s<t\}$相互独立，实系数$\varphi_1, \cdots, \varphi_p, \theta_1, \cdots, \theta_q$使得多项式

$$\varphi(z) = 1 - \varphi_1 z - \cdots - \varphi_p z^p \neq 0, |z| \leqslant 1, \quad (1.2.5)$$

$$\theta(z) = 1 - \theta_1 z - \cdots - \theta_q z^q \neq 0, |z| \leqslant 1, \quad (1.2.6)$$

且$\varphi(z)$与$\theta(z)$互质，$\varphi_p\theta_q \neq 0$，则称$\{x_t\}$为服从(p,q)阶的线性自回归滑动平均模型，记为LARMA(p,q)模型，p表示(1.2.4)式中自回归部分的阶数，q表示滑动平均部分的阶数. 也称$\{x_t\}$为LARMA(p,q)序列.

关于LARMA(p,q)序列有如下的表示方式和性质：

定理 1.2.3 设$\{x_t\}$是满足(1.2.4)式的LARMA(p,q)序列，则有

(1) $\{x_t\}$是模型(1.2.4)的唯一的平稳解，且

$$x_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad \psi_0 = 1, \quad (1.2.7)$$

其中系数序列$\{\psi_j\}$满足

$$\left(\sum_{j=0}^{\infty} \psi_j z^j\right) \varphi(z) = \theta(z), \quad (1.2.8)$$

并存在常数$c>0$和$0<\rho<1$，使得

$$|\psi_j| \leqslant c\rho^j \quad (j \geqslant 0); \quad (1.2.9)$$

(2) $\{x_t\}$是宽平稳序列，且有谱密度

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \frac{\theta(e^{i\lambda})}{\varphi(e^{i\lambda})} \right|^2; \quad (1.2.10)$$

(3) ε_t可唯一地表为

$$\varepsilon_t = \sum_{j=0}^{\infty} \pi_j x_{t-j}, \pi_0 = 1, \quad (1.2.11)$$

其中系数序列$\{\pi_j\}$满足

$$\left(\sum_{j=0}^{\infty} \pi_j z^j\right) \theta(z) = \varphi(z), \quad (1.2.12)$$

并存在常数$c'>0$和$0<\rho'<1$，使得

$$|\pi_j| \leqslant c'(\rho')^j \quad (j \geqslant 0). \tag{1.2.13}$$

由以上性质可知，LARMA 序列是一类特殊的线性序列. 关于 LARMA 序列的研究，已有非常丰富的文献可参考(如文献[16]、[58]等).

迄今为止，LARMA 模型不仅是时间序列分析中研究最多的，也是应用最为广泛的. 其原因在于，研究 LARMA 模型的概率结构、统计的理论和方法，有比较成熟的数学和概率论工具可使用. 从计算角度考虑，根据实际数据建立 LARMA 模型的统计方法，也比较容易实现，并有许多软件可供使用. 因此，LARMA 模型在许多领域被普遍使用.

但是，由于 LARMA 序列是一种线性序列，关于它的结构特征和建模方法的研究，主要是与其自协方差函数有关，因此，用 LARMA 模型的研究方法难于揭示二阶矩结构以外的结构信息. 而在实际应用中，又不断地提出分析实际数据中存在着比线性结构更丰富的信息的要求. 这就刺激了对非线性时间序列的研究.

§1.3 非线性时间序列

简单地说，凡是不能表示成(1.2.1)式的线性序列形式的时间序列，都是非线性时间序列. 这样定义非线性时间序列太泛泛了. 实际上，在非线性时间序列的分析中，我们主要研究平稳非线性序列.

对于非线性时间序列的研究，大致可分为两种方法. 一种是研究一般平稳非线性序列. 它又有时域方法和频域方法之分. 以下分别举例说明之.

高阶矩描述方法就是一种时域方法. 它是通过引进序列的高阶矩来刻划其结构特征的. 例如，引入三阶矩函数：

$$\gamma_{k,j} = E(x_t - \mu)(x_{t+k} - \mu)(x_{t+j} - \mu) \quad (k, j = 0, \pm 1, \cdots), \tag{1.3.1}$$

其中 $\mu = Ex_t$. 连同序列的自协方差函数 γ_k，则描述了序列一至三阶矩的相依结构. 类似地，可以考虑比三阶矩更高的矩函数. 在实际中，一般只考虑到有限阶的矩函数. 因此，它们都只是序列的部分结构信息. 况且，由于计算的困难，也不能考虑太高的矩函数.

高阶谱描述方法是与高阶矩描述方法等价的，是一种频域方法. 它是借助于高阶矩的谱表示，引进高阶谱的概念. 例如，与三阶矩函数相应地有如下表达式：

$$\gamma_{k,j}=\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}(2\pi)^{-2}\exp\{ik\lambda+ij\omega\}F(\mathrm{d}\lambda,\mathrm{d}\omega),\qquad(1.3.2)$$

其中 $F(\lambda,\omega)$为序列的双谱. 正如平稳序列的谱函数与其自协方差函数有相互对应关系(1.1.2)一样,(1.3.2)式表达了三阶矩函数与双谱之间的关系. 本书将不详细介绍这一描述方法. 读者可参阅文献[19].

Volterra 展开方法是将平稳序列$\{x_t\}$展成如下形式的求和项:

$$\begin{aligned}x_t=a&+\sum_{j=0}^{\infty}a_j\varepsilon_{t-j}+\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}a_{ij}\varepsilon_{t-i}\varepsilon_{t-j}\\&+\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}a_{ijk}\varepsilon_{t-i}\varepsilon_{t-j}\varepsilon_{t-k}+\cdots,\end{aligned}\qquad(1.3.3)$$

式中$\{\varepsilon_t\}$为白噪声序列. 它也属于时域方法. 显然,(1.3.3)式是线性序列(1.2.1)的直接推广. 但是,无论是在理论研究方面还是在应用方面,都只能考虑(1.3.3)式中的前有限项求和,以此做为 x_t 的近似展开. 与前述的高阶谱和高阶矩方法相似,这种近似展开也只能描述序列的部分结构信息. 而且,即使考虑到三阶项 $a_{ijk}\varepsilon_{t-i}\varepsilon_{t-j}\varepsilon_{t-k}$的求和项为止,在实际应用中也是非常难以处理的. 本书也不详细介绍此方法.

研究非线性时间序列的另一种方法是模型分析法. 模型的形式是多种多样的,我们先讨论熟知的 LARMA 模型的推广模型.

考虑如下非线性 ARMA(p,q)模型(Nonlinear ARMA(p,q))

$$x_t=\psi(x_{t-1},\cdots,x_{t-p},\varepsilon_t,\varepsilon_{t-1},\cdots,\varepsilon_{t-q}),\qquad(1.3.4)$$

其中$\{\varepsilon_t\}$为白噪声序列,且 ε_t 与$\{x_s,s<t\}$相互独立. 模型(1.3.4)简记为 NLARMA (p,q)模型. 必须注意,(1.3.4)式做为(1.2.4)式所描述的 LARMA(p,q)模型的推广有两种情况. 一种情况是:(1.3.4)式中的函数 ψ 为给定的非线性函数,且仅依赖于有限个参数. 我们称这类模型为**非线性参数模型**,例如,

$$x_t=\alpha x_{t-1}+\beta x_{t-1}\varepsilon_{t-1}+\varepsilon_t,\qquad(1.3.5)$$

其中$\{\varepsilon_t\}$为白噪声序列,α 和 β 为未知参数. 模型(1.3.5)属于本书将要讨论的一类非线性时间序列模型——双线性模型. 另一种情况是(1.3.4)式中的函数 ψ,是非参数型的非线性函数. 注意,(1.2.4)式中的平稳序列$\{x_t\}$不仅是线性的,而且模型还是有限参数模型. 但是,NLARMA 模型(1.3.4)既可为参数模型,也可为非参数模型. 对一般的 NLARMA 模型(1.3.4)的理论研究是非常困难的,文献中代之以研究模型(1.3.4)的一些特殊类型.

首先,(1.3.4)式的一种最简单也是最重要的特殊类型是如下的非线性自回归(Nonlinear Autoregression)模型:

$$x_t = \varphi(x_{t-1},\cdots,x_{t-p}) + \varepsilon_t, \tag{1.3.6}$$

其中φ是R^p到R^1的可测函数,$\{\varepsilon_t\}$是白噪声序列,满足条件:

$$E\varepsilon_t = 0, E\varepsilon_t^2 = \sigma^2 < \infty,\ \varepsilon_t 与 \{x_s, s < t\} 相互独立. \tag{1.3.7}$$

p称为模型(1.3.6)的阶数.模型(1.3.6)简记为NLAR(p)模型.

模型(1.3.6)的特殊情形是如下线性自回归(Linear Autoregression)模型:

$$x_t = \varphi_1 x_{t-1} + \cdots + \varphi_p x_{t-p} + \varepsilon_t, \tag{1.3.8}$$

式中$\{\varepsilon_t\}$为白噪声序列,满足条件(1.3.7)式,实系数$\varphi_1,\cdots,\varphi_p$满足条件(1.2.5)式.模型(1.3.8)简记为LAR(p)模型.注意,(1.3.8)式是(1.2.4)式中$q=0$的特殊情况.而(1.3.6)式作为(1.3.8)式的推广还包含了向非参数模型的推广.在文献中,为了更深入地研究NLAR模型(1.3.6),主要是对非线性函数φ的各种特殊类型分别进行研究.这也是本书的主要研究内容.

比(1.3.6)式更复杂一点的非线性模型是

$$x_t = \varphi(x_{t-1},\cdots,x_{t-p}) + \varepsilon_t s(x_{t-1},\cdots,x_{t-p}), \tag{1.3.9}$$

式中φ和s都是R^p到R^1的可测函数,且s是处处为正的.$\{\varepsilon_t\}$是白噪声序列,满足条件(1.3.7)式.显然,当$s(x_{t-1},\cdots,x_{t-p})$等于常数时,(1.3.9)式便退化为(1.3.6)式.特别当$\varphi(x_{t-1},\cdots,x_{t-p})$为线性函数时,模型(1.3.9)与LAR($p$)模型(1.3.8)有着更密切的联系.

再比(1.3.9)式更复杂些的非线性模型是

$$x_t = \varphi(x_{t-1},\cdots,x_{t-p}) + s(x_{t-1},\cdots,x_{t-p},\varepsilon_t), \tag{1.3.10}$$

其中s是R^{p+1}到R^1的可测函数,φ和$\{\varepsilon_t\}$如(1.3.9)式所假定.或更一般地为

$$x_t = \varphi(x_{t-1},\cdots,x_{t-p},\varepsilon_t), \tag{1.3.11}$$

其中φ是R^{p+1}到R^1的可测函数,$\{\varepsilon_t\}$的假定同(1.3.9)式.

显然,(1.3.11)式为LAR(p)模型(1.3.8)的最一般的推广.一般说来,越一般的模型越难深入研究.研究形如(1.3.11)式的NLAR模型已经非常困难了,本书将很少讨论此种模型.

在研究LARMA模型时,很容易验证(1.2.4)式有平稳解的条件,即(1.2.5)式,并给出(1.2.4)式的平稳解的表达式(1.2.7)式.但是,在研究NLARMA模型(1.3.4)时,讨论其有平稳解的条件,比线性情况要困难

得多,所使用的理论与方法也有很多实质性的差异.在以后的章节中,我们将看到,一般状态空间马尔科夫链(以下简称马氏链)的理论与方法,在研究形如(1.3.6)、(1.3.9)等式的 NLAR 模型的平稳解、遍历性时起着重要的作用.

在 LARMA 模型中,当 $p=0$ 时,(1.2.4)式成为 q 阶线性滑动平均(Linear Moving Average)模型,简记为 LMA(q)模型,即

$$x_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q}. \tag{1.3.12}$$

易见由(1.3.12)式所确定的$\{x_t\}$是平稳序列.此模型的非线性推广为

$$x_t = \psi(\varepsilon_t, \varepsilon_{t-1}, \cdots, \varepsilon_{t-q}), \tag{1.3.13}$$

其中 ψ 为 R^{q+1}到 R^1 的可测函数,$\{\varepsilon_t\}$为白噪声序列.它同样既可为参数模型,也可为非参数模型.显然,由(1.3.13)式所确定的$\{x_t\}$是平稳序列.它与 Volterra 展式(1.3.3)相似,在研究其平稳性、遍历性方面并无困难.但是,对其进行统计分析却很困难,远不如 LMA(q)模型(1.3.12)那样容易.

§1.4　平稳序列的预报及新息序列的假定

在本节里,我们从预报的观点分析时间序列的线性性与非线性性,并对预报误差序列的各种不同假定加以讨论.为此,我们考虑具有有穷方差的平稳序列,依性质 1.1.4 知其还是宽平稳序列.由于此类序列的均值为常数,故不妨只讨论零均值的情况.

设$\{x_t\}$为平稳序列,$Ex_t=0$,$Ex_t^2<\infty$.根据时间序列$\{x_t\}$的历史观测值 $x_n, x_{n-1}, \cdots$,对未来时刻 $n+k(k>0)$时的取值 x_{n+k}进行估计,称之为**时间序列的预报**,所得的估计值记为 $\hat{x}_{n+k}$,称为 x_{n+k}的**预报值**或$\{x_t\}$的 **k 步预报值**.当 $k=1$ 时,称为**一步预报**.文献中有许多不同的预报方法,我们在这里主要讨论两种预报方法:一种是条件期望的预报方法,即对给定的 $x_n, x_{n-1}, \cdots$,以 x_{n+1}的条件均值

$$\hat{x}_n(1) = E\{x_{n+1} | x_n, x_{n-1}, \cdots\} \tag{1.4.1}$$

作为 x_{n+1}的预报.

另一种是线性投影方法,即寻找以下极值问题的极小化解:

$$\inf_{c_1, c_2, \cdots} E\left\{x_{n+1} - \sum_{j=1}^{\infty} c_j x_{n+1-j}\right\}^2. \tag{1.4.2}$$

为叙述方便,将(1.4.2)式的极小化解表示为

$$\hat{x}_n^*(1) = \hat{E}\{x_{n+1} | x_n, x_{n-1}, \cdots\}. \tag{1.4.3}$$

以下随机序列称为*预报误差序列*：

$$e_t = x_t - \hat{x}_{t-1}(1), \tag{1.4.4}$$

$$e_t^* = x_t - \hat{x}_{t-1}^*(1). \tag{1.4.5}$$

在文献中，预报误差序列$\{e_t\}$和$\{e_t^*\}$又称为*新息序列*. 为了区别(1.4.4)和(1.4.5)两式所确定的预报误差序列，我们称$\{e_t^*\}$为*投影新息序列*. 以下性质在时间序列分析中是熟知的.

性质 1.4.1 设$\{x_t\}$是平稳序列，$Ex_t=0, Ex_t^2<\infty$，则有

(1) $\{e_t\}$和$\{e_t^*\}$都是零均值的平稳序列，具有有限方差；

(2) $E\{e_t | x_{t-1}, x_{t-2}, \cdots\}=0$, a. s.；

(3) 对任何$k\geqslant 1$和实数c_j $(j=1,2,\cdots,k)$，有

$$E\left\{e_t^* \cdot \sum_{j=1}^{k} c_j x_{t-j}\right\} = 0;$$

(4) $E\{x_t-\hat{x}_{t-1}(1)\}^2 \leqslant E\{x_t-\hat{x}_{t-1}^*(1)\}^2$；

(5) $E\{x_t-\hat{x}_{t-1}(1)\}^2 \leqslant E\{x_t-\xi(x_{t-1}, x_{t-2}, \cdots)\}^2$，

其中ξ是任意可测函数.

性质 1.4.1 中的(2)表明$\{e_t\}$是鞅差序列. (3)表明e_t^*与$x_{t-1}, x_{t-2}, \cdots$的任何线性组合都不相关. (4)表明条件均值预报比投影预报的误差方差小，也可以说，前者比后者好. (5)表明条件均值预报的误差方差比任何其他预报的误差方差都小. 如果以预报误差方差的大小作为评判预报优劣的准则，那么条件均值预报是最好的预报.

利用平稳序列的预报概念，可以给出平稳序列线性性的另一种定义，即当$\hat{x}_{t-1}(1)=\hat{x}_{t-1}^*(1)$时，称相应的平稳序列为*预报线性序列*. 一般说来，平稳序列的两种预报并不相同，即$\hat{x}_{t-1}(1)\neq\hat{x}_{t-1}^*(1)$. 但是，容易验证，满足(1.2.1)式的线性序列必有$\hat{x}_{t-1}(1)=\hat{x}_{t-1}^*(1)$. 从而根据(1.2.7)式知，LARMA 序列也是预报线性序列. 此外，不难验证，平稳正态序列也是预报线性序列.

现举一例说明$\hat{x}_{t-1}(1)\neq\hat{x}_{t-1}^*(1)$是可能的.

例 1.4.2 考虑如下平稳随机序列：

$$x_t = \varepsilon_t + \frac{1}{9}\varepsilon_{t-1}\varepsilon_{t-2}, \tag{1.4.6}$$

其中$\{\varepsilon_t\}$是正态白噪声序列，$E\varepsilon_t=0, E\varepsilon_t^2=1$. 易见，(1.4.6)式所确定的平稳序列$\{x_t\}$满足

$$Ex_t = E\left(\varepsilon_t + \frac{1}{9}\varepsilon_{t-1}\varepsilon_{t-2}\right) = 0,$$

$$Ex_t^2 = E\left(\varepsilon_t + \frac{1}{9}\varepsilon_{t-1}\varepsilon_{t-2}\right)^2 = \frac{82}{81}.$$

从而$\{x_t\}$是宽平稳的.另外,又易见,对$k\geqslant 1$,

$$Ex_tx_{t+k} = E\left(\varepsilon_t + \frac{1}{9}\varepsilon_{t-1}\varepsilon_{t-2}\right) \times \left(\varepsilon_{t+k} + \frac{1}{9}\varepsilon_{t+k-1}\varepsilon_{t+k-2}\right) = 0,$$

故

$$\hat{x}_{t-1}^*(1) = 0,$$

$$E\{x_t - \hat{x}_{t-1}^*(1)\}^2 = Ex_t^2 = \frac{82}{81}.$$

为了说明$\hat{x}_{t-1}(1)\neq\hat{x}_{t-1}^*(1)$,我们不一定非要求出$\hat{x}_{t-1}(1)$来,只要注意到

$$\begin{aligned}E\left\{x_t - \frac{1}{9}x_{t-1}x_{t-2}\right\}^2 &= E\left\{\varepsilon_t + \frac{1}{9}\varepsilon_{t-1}\varepsilon_{t-2} - \frac{1}{9}\left(\varepsilon_{t-1} + \frac{1}{9}\varepsilon_{t-2}\varepsilon_{t-3}\right)\right.\\ &\quad\left.\times\left(\varepsilon_{t-2} + \frac{1}{9}\varepsilon_{t-3}\varepsilon_{t-4}\right)\right\}^2\\ &= 1 + 4\left(\frac{1}{81}\right)^2 + 3\left(\frac{1}{729}\right)^2 < \frac{82}{81}\\ &= E\{x_t - \hat{x}_{t-1}^*(1)\}^2.\end{aligned}$$

利用性质1.4.1中的(4)和(5),有

$$\begin{aligned}E\{x_t - \hat{x}_{t-1}(1)\}^2 &\leqslant E\left\{x_t - \frac{1}{9}x_{t-1}x_{t-2}\right\}^2\\ &< E\{x_t - \hat{x}_{t-1}^*(1)\}^2.\end{aligned}$$

从而可知

$$\hat{x}_{t-1}(1) \neq \hat{x}_{t-1}^*(1).$$

如若不然,必有

$$E\{x_t - \hat{x}_t(1)\}^2 = E\{x_t - \hat{x}_{t-1}^*(1)\}^2,$$

此式与前式相矛盾!

现在讨论平稳序列的新息序列$\{e_t\}$的方差的性质.根据性质1.4.1中的(2)知,若以条件期望方法给出平稳序列的一步预报,则预报误差序列$\{e_t\}$是鞅差序列,且是平稳的.根据性质1.4.1中的(4)和(5)知,误差e_t的方差具有相对较小的性质.但是,预报误差e_t的条件方差

$$\sigma_{t-1}^2(1) = E\{e_t^2 | x_{t-1}, x_{t-2}, \cdots\} \tag{1.4.7}$$

更精细地刻划了一步预报误差方差与平稳序列的历史值间的联系. 一般说来,条件方差 $\sigma_{t-1}^2(1)$是 $x_{t-1}, x_{t-2}\cdots$的可测函数. 只在某些特殊情况下,$\sigma_{t-1}^2(1)$才取常数值,而与 $x_{t-1}, x_{t-2}, \cdots$无关,即

$$E\{e_t^2 | x_{t-1}, x_{t-2}, \cdots\} = \sigma^2. \tag{1.4.8}$$

根据以上的讨论,新息序列$\{e_t\}$被假定为平稳鞅差序列是最合理的. 但是,在时间序列分析的较早的文献中,$\{e_t\}$被假定为白噪声序列. 在较近代的文献中,又被较弱的条件所代替,即假定$\{e_t\}$为平稳鞅差序列,并且(1.4.8)式成立. 这些关于新息序列的假定都较前述的平稳鞅差序列的假定带有较强的限制. 这些较强的假定,对时间序列分析的理论研究带来了很多方便. 但是,它们却缩小了许多有用模型的应用背景.

在研究非线性时间序列时,我们发现,上述关于新息序列的不同假定,不仅在理论与应用方面起着不同的作用,而且不同的新息序列的假定还导致不同的非线性序列的模型结构,也使时间序列的线性性有不同的含义. 这些内容在本书中将进行详细讨论.

最后,我们简单讨论一下投影新息序列$\{e_t^*\}$的正交性质. 根据投影性质或性质 1.4.1 中的(1)和(3)可知

$$Ee_t^* = 0, \ Ee_t^* e_s^* = 0 \quad (t \neq s). \tag{1.4.9}$$

从而$\{e_t^*\}$为正交序列. 或者说,对于 $s \neq t$,e_t^* 与 e_s^* 是一对不相关的随机变量. 显然,新息序列$\{e_t\}$在前面几种不同的假定下,都是正交序列. 也就是说,把预报误差序列假定为正交序列属更弱的假定. 对于平稳正态序列,根据正态随机变量的独立性与正交性的等价性,其新息序列总是白噪声序列. 所以,在平稳正态序列情况下,以上所提到的各种不同的有关新息序列的假定都是等价的.

§ 1.5 线性与非线性自回归模型

在 § 1.4 中,如果(1.4.1)式的一步预报只依赖于有限的历史值 $x_{t-1}, \cdots, x_{t-p}$,即

$$\begin{aligned} \hat{x}_{t-1}(1) &= E\{x_t | x_{t-1}, x_{t-2}, \cdots\} \\ &= E\{x_t | x_{t-1}, \cdots, x_{t-p}\} \\ &\equiv \varphi(x_{t-1}, \cdots, x_{t-p}). \end{aligned} \tag{1.5.1}$$

联合(1.4.4)式,有

$$x_t = \varphi(x_{t-1}, \cdots, x_{t-p}) + e_t, \tag{1.5.2}$$

其中新息序列$\{e_t\}$为平稳鞅差序列.(1.5.2)式在形式上与NLAR(p)模型(1.3.6)式相同.但在(1.3.6)式中,$\{e_t\}$被假定为白噪声序列.这种假定的不同将导致不同的模型结构,以及不同的非线性自回归模型的定义.详细讨论将在第3、4章中进行.在这一节里,我们仍旨在浅释线性与非线性时间序列的异同.为此,以下暂时只讨论$\{e_t\}$为白噪声序列的情况,并且在这里我们仅讨论一阶NLAR模型,即

$$x_t = \varphi(x_{t-1}) + \varepsilon_t. \tag{1.5.3}$$

当(1.5.3)式中的函数φ为线性函数

$$x_t = \varphi x_{t-1} + \varepsilon_t \tag{1.5.4}$$

时,相应的平稳可逆性条件为

$$|\varphi| < 1. \tag{1.5.5}$$

反复使用(1.5.4)式的迭代关系,可得到

$$\begin{aligned} x_t &= \varepsilon_t + \varphi x_{t-1} = \varepsilon_t + \varphi\varepsilon_{t-1} + \varphi^2 x_{t-2} \\ &= \cdots = \varepsilon_t + \varphi\varepsilon_{t-1} + \cdots + \varphi^n\varepsilon_{t-n} + \varphi^{n+1}x_{t-n-1}. \end{aligned} \tag{1.5.6}$$

利用条件(1.5.5)有

$$\varphi^{n+1}x_{t-n-1} \to 0 \quad (\text{当 } n \to \infty),$$

$$\varepsilon_t + \varphi\varepsilon_{t-1} + \cdots + \varphi^n\varepsilon_{t-n} \to \sum_{j=0}^{\infty}\varphi^j\varepsilon_{t-j} \quad (\text{当 } n \to \infty).$$

以上两式的收敛以概率为1地成立.从而,由(1.5.6)式得到

$$x_t = \sum_{j=0}^{\infty}\varphi^j\varepsilon_{t-j}. \tag{1.5.7}$$

上式给出了模型(1.5.4)的平稳解的表达式.

但是,讨论非线性自回归模型(1.5.3)的平稳性条件以及求其平稳解,要比线性自回归模型(1.5.4)困难得多.在假定$\{\varepsilon_t\}$为白噪声序列时,从递推关系式(1.5.6)不难看出,(1.5.4)式有平稳解的充分必要条件是(1.5.5)式成立.对非线性自回归模型(1.5.3)迄今还未得到其有平稳解的充分必要条件.即使只考虑其充分条件,也并不容易.为了说明非线性自回归模型和线性自回归模型的联系,我们要讨论一种较容易验证的充分条件.

首先注意,从(1.5.4)式到(1.5.6)式,又从(1.5.6)式得到(1.5.7)式的推导方法,可以称为向后递推法.这一方法在非线性情况下并不是总能仿效的.这是因为,利用形如(1.5.3)式的递推关系式不能得出类似于

(1.5.6)式的显式递推关系式，而需要代之以隐式表达式：

$$\begin{aligned} x_t &= \varphi(x_{t-1}) + \varepsilon_t \\ &= \varphi(\varphi(x_{t-2}) + \varepsilon_{t-1}) + \varepsilon_t \\ &= \cdots \\ &= \varphi(\varphi(\varphi(\cdots\varphi(\varphi(\varphi(x_{t-n}) \\ &\qquad + \varepsilon_{t-n+1}) + \varepsilon_{t-n+2})\cdots) + \varepsilon_{t-2}) + \varepsilon_{t-1}) + \varepsilon_t. \end{aligned} \tag{1.5.8}$$

我们很难找到(1.5.8)有极限的充分必要条件，更难找到极限的表达式. 以下我们来证明：如果假定非线性函数 $\varphi(\cdot)$满足

$$\begin{aligned} &\varphi(0) = 0, \\ &|\varphi(x) - \varphi(y)| \leqslant \rho|x - y| \quad (\text{对一切 } x, y \in R^1), \end{aligned} \tag{1.5.9}$$

其中实数 ρ 满足 $0<\rho<1$，则(1.5.3)式有平稳解.

在(1.5.8)式中用零值代替 x_{t-n}，并记为：

$$\begin{aligned} \xi_{n,t} = &\varphi(\varphi(\cdots\varphi(\varphi(\varphi(0) \\ &+ \varepsilon_{t-n+1}) + \varepsilon_{t-n+2})\cdots) + \varepsilon_{t-1}) + \varepsilon_t, \end{aligned} \tag{1.5.10}$$

这样我们便得到一串随机变量序列$\{\xi_{n,t}\}$. 于是，要证明在条件(1.5.9)下，(1.5.3)式有平稳解，只需证明在条件(1.5.9)下，当 $n\to\infty$时，$\xi_{n,t}$有极限，且此极限恰好是(1.5.3)式的解. 为此引入如下记号：

$$\xi_{n,t} \triangleq \varphi(0, \varepsilon_{t-n+1}, \varepsilon_{t-n+2}, \cdots, \varepsilon_t).$$

由(1.5.10)式知，上式右端表示从 $x_{t-n}=0$ 出发，利用(1.5.3)式向后递推 n 次后的值. 设 $n>m$，利用(1.5.9)式和(1.5.10)式，有

$$\begin{aligned} |\xi_{n,t} - \xi_{m,t}| &= |\varphi(\xi_{n-1,t-1}) + \varepsilon_t - \varphi(\xi_{m-1,t-1}) - \varepsilon_t| \\ &\leqslant \rho|\xi_{n-1,t-1} - \xi_{m-1,t-1}|. \end{aligned}$$

反复利用(1.5.9)式，完全相同于以上推导，得

$$\begin{aligned} |\xi_{n,t} - \xi_{m,t}| &\leqslant \rho|\xi_{n-1,t-1} - \xi_{m-1,t-1}| \\ &\leqslant \cdots \\ &\leqslant \rho^{m-1}|\xi_{n-m+1,t-m+1} - \xi_{1,t-m+1}| \\ &\leqslant \rho^{m-1}|\varphi(\xi_{n-m,t-m}) + \varepsilon_{t-m+1} - \varepsilon_{t-m+1}| \\ &= \rho^{m-1}|\varphi(\xi_{n-m,t-m}) - \varphi(0)| \\ &\leqslant \rho^m|\xi_{n-m,t-m}| \\ &= \rho^m|\varphi(\xi_{n-m-1,t-m-1}) + \varepsilon_{t-m}| \\ &\leqslant \rho^m\{|\varepsilon_{t-m}| + \rho|\xi_{n-m-1,t-m-1}|\} \\ &\leqslant \rho^m\{|\varepsilon_{t-m}| + \rho|\varepsilon_{t-m-1}| + \cdots + \rho^{n-m}|\varepsilon_{t-n+1}|\} \end{aligned}$$

$$\leqslant \rho^m \sum_{j=0} \rho^j |\varepsilon_{t-m-j}| \to 0 \quad (\text{当 } m \to \infty),$$

上式以概率为 1 地成立. 从而$\{\xi_{n,t}\}$是柯西序列. 于是$\{\xi_{n,t}\}$以概率为 1 地有极限,记为

$$x_t = \lim_{n\to\infty} \xi_{n,t}. \tag{1.5.11}$$

由(1.5.3)式和(1.5.10)式知:

$$\xi_{n,t} = \varphi(\xi_{n-1,t-1}) + \varepsilon_t.$$

再由(1.5.11)式有

$$x_t = \varphi(x_{t-1}) + \varepsilon_t.$$

这就意味着$\{x_t\}$是(1.5.3)的解. 再由$\{\xi_{n,t}\}$的定义和$\{\varepsilon_t\}$为白噪声序列,又可证明$\{x_t\}$为平稳序列.

容易验证(1.5.4)式中的线性函数$\varphi(x)=\varphi x$满足(1.5.9)式. 通过上述讨论可以看出,在条件(1.5.9)下证明(1.5.3)式有平稳解要比线性情况麻烦得多,而且未能给出解的表达式. 虽然上述向后递推方法不难推广到多元时间序列情况,但是,条件(1.5.9)式实属较强的限制,至少它把所有不连续函数φ排除在外. 具有分片连续函数的非线性时间序列模型是实际数据分析中很有用的一类模型,例如第 3 章中将要讨论的门限自回归模型就是这一类模型. 上述方法不能推广到此类非线性时间序列模型. 因此,在研究非线性自回归模型时,并不着眼于上述的向后递推方法,而是借助于一般状态空间马氏链的理论与方法. 这将在第 4 章中详细讨论.

从线性自回归模型到非线性自回归模型,在讨论上述的推移方法时,还可以从另一角度展示它们的本质差别. 在讨论 LAR(1)模型(1.5.4)时,它有平稳解的充分必要条件是(1.5.5)式成立. 而这又是以下线性差分方程

$$y_t = \varphi y_{t-1} \quad (t \geqslant 1) \tag{1.5.12}$$

在零点稳定的充分必要条件. 在讨论 NLAR(1)模型(1.5.3)时,它有平稳解的一个充分条件是(1.5.9)式成立. 这也是以下非线性差分方程

$$y_t = \varphi(y_{t-1}) \quad (t \geqslant 1) \tag{1.5.13}$$

在零点稳定的充分条件. 要给出非线性差分方程(1.5.13)在零点稳定的充分必要条件是非常困难的. 由此也可看出非线性自回归模型和线性自回归模型的差别之大. 虽然我们只以一阶模型为例说明上述的差别,但对于一般的非线性自回归模型与线性自回归模型也有类似的差别. 实际上,NLAR(p)模型(1.5.2)有平稳解的条件与相应的p阶非线性差分方程

$$y_t = \varphi(y_{t-1}, \cdots, y_{t-p}) \tag{1.5.14}$$

密切相关.当(1.5.14)式为线性差分方程时,其稳定性的条件不难获得.对于非线性情况,则非常困难.关于形如(1.5.14)式的差分方程的研究,与当今比较活跃的混沌研究有密切关系.为此,我们将在下一章中概要地介绍由(1.5.4)式确定的离散动态系统的有关概念性知识.这些内容将作为研究非线性时间序列模型的概率结构的预备知识.

第 2 章

离散动态系统概述

在时间序列分析中，自回归模型既是一种带有噪声项的离散动态系统，又是一种带有噪声项的随机差分方程. 若将噪声项用零值代替，则导出相应的离散动态系统. 实质上，这是一种非随机的差分方程，而实际中的差分方程几乎又总难免介入或大或小的噪声干扰项. 由此可见，自回归模型与离散动态系统有着密切的联系. 线性自回归模型对应着线性离散动态系统，其结构特征较为简单，而非线性自回归模型则对应着非线性离散动态系统，其结构特征的复杂程度是难以想象的. 到目前为止，关于非线性动态系统的研究远未达到成熟的程度. 而讨论非线性时间序列又不能不涉及非线性离散动态系统的概念. 因此，本章将介绍有关的离散动态系统的概念和基本理论，作为以后章节的预备知识.

§ 2.1 线性离散动态系统

本节将简要介绍研究线性离散动态系统的主要方法和结果. 为行文方便，以后提及线性(非线性)系统时均指线性(非线性)离散动态系统. 我们先讨论一阶一维线性系统，再推广到一阶高维线性系统. 而高阶一维线性系统可转化为一阶高维线性系统.

定义 2.1.1 以下的迭代关系称为一阶一维线性系统，即

$$\begin{cases} x_t = a_1 x_{t-1} + a_0 \quad (t \geqslant 1); \\ x_0 \in R^1, \end{cases} \tag{2.1.1}$$

其中 a_0 和 a_1 为实参数.

线性系统(2.1.1)的迭代序列 $\{x_t, t \geqslant 0\}$ 是由参数 a_0 和 a_1 及初始值 x_0 唯一确定的. 它具有如下表达式：

$$\begin{aligned} x_t &= a_0 + a_1 x_{t-1} = a_0 + a_1(a_0 + a_1 x_{t-2}) \\ &= a_0 + a_0 a_1 + a_1^2 x_{t-2} = \cdots \\ &= a_0 + a_0 a_1 + a_0 a_1^2 + \cdots + a_0 a_1^{t-1} + a_1^t x_0. \end{aligned} \tag{2.1.2}$$

讨论由(2.1.1)式产生的迭代序列$\{x_t,t\geqslant 0\}$的极限性质,是线性系统(2.1.1)的结构特征的重要描述方法.容易验证如下性质:

性质 2.1.2 对由(2.1.1)式产生的迭代序列$\{x_t,t\geqslant 0\}$,有

(i) 当$|a_1|<1$时,对任何$x_0\in R^1$,有

$$\lim_{t\to\infty} x_t = \frac{a_0}{1-a_1};$$

(ii) 当$|a_1|>1$时,对任何$x_0\in R^1$,且$x_0\neq a_0/(1-a_1)$,有

$$\lim_{t\to\infty}|x_t| = \infty,$$

且当$x_0=a_0/(1-a_1)$时,有

$$x_t = x_0 = \frac{a_0}{1-a_1};$$

(iii) 当$a_1=1$,且$a_0\neq 0$时,对任何$x_0\in R^1$,有

$$\lim_{t\to\infty}|x_t| = \infty;$$

(iv) 当$a_1=1,a_0=0$时,对任何$t\geqslant 0$和$x_0\in R^1$,有

$$x_t = x_0;$$

(v) 当$a_1=-1$,对任何$x_0\in R^1$,如果$x_0\neq a_0/2$,则对任何$k\geqslant 0$,有

$$x_{2k+1} = a_0 - x_0, \qquad x_{2k+2} = x_0;$$

如果$x_0=a_0/2$,则对任何$t\geqslant 0$,有

$$x_t = x_0.$$

利用性质 2.1.2,我们可对线性系统(2.1.1)作如下初步分类:

定义 2.1.3 对线性系统(2.1.1),

(i) 当$|a_1|<1$时,称其为稳定的线性系统;

(ii) 当$|a_1|>1$且$a_0\neq 0$,或$a_1=1$且$a_0\neq 0$时,称其为发散的线性系统;

(iii) 当$a_1=1$且$a_0=0$时,称其为不变线性系统;

(iv) 当$a_1=-1$时,称其为非稳定的周期振荡系统.

将以上讨论推广到一阶高维线性系统时,是没有实质性困难的.所不同的是,必须利用矩阵和向量运算.又由于高维向量中的不同分量可以属于不同的分类,使分类情况更为复杂.本节的目的仅在于通过回顾线性系统的理论,以便于对比和了解非线性系统中的新概念和新的描述方法.因此,我们不去仔细地讨论向量情形的分类问题,而仅给出其定义和稳定性等概念.

定义 2.1.4 以下迭代关系称为一阶m维线性系统:

$$\begin{cases} \boldsymbol{x}_t = A\boldsymbol{x}_{t-1} + \boldsymbol{a} \quad (t \geqslant 1); \\ \boldsymbol{x}_0 \in R^m, \end{cases} \tag{2.1.3}$$

其中 A 为 m 阶方阵，$\boldsymbol{a}$ 为 m 维向量.

一阶一维线性系统(2.1.1)的另一种推广是 p 阶一维线性系统.

定义 2.1.5 以下迭代关系称为 p 阶一维线性系统：

$$\begin{cases} x_t = a_0 + a_1 x_{t-1} + \cdots + a_p x_{t-p} \quad (t \geqslant 1); \\ x_k \in R^1 \quad (k = 0, -1, \cdots, -p+1), \end{cases} \tag{2.1.4}$$

其中 $a_0, a_1, \cdots, a_p$ 为实系数；$x_0, x_{-1}, \cdots, x_{-p+1}$ 为初始值.

令

$$\boldsymbol{x}_t = (x_t, x_{t-1}, \cdots, x_{t-p+1})^\tau,$$

$$\boldsymbol{a} = (a_0, 0, \cdots, 0)^\tau,$$

$$A = \begin{pmatrix} a_1 & a_2 & \cdots & a_{p-1} & a_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}. \tag{2.1.5}$$

于是(2.1.4)式等价于如下的一阶 p 维线性系统：

$$\begin{cases} \boldsymbol{x}_t = A\boldsymbol{x}_{t-1} + \boldsymbol{a} \quad (t \geqslant 1); \\ \boldsymbol{x}_0 \in R^p. \end{cases} \tag{2.1.6}$$

比较(2.1.6)式与(2.1.3)式，其表示完全相同，而差别在于前者的 $\boldsymbol{a}$ 和方阵 A 具有一定的特殊形式. 总之，研究 p 阶一维线性系统，可转化为研究一阶 p 维线性系统.

依照完全类似的技巧，一个 p 阶 m 维线性系统，可化为一个等价的一阶 pm 维线性系统. 总之，任意有穷阶有限维的线性系统都可转化为一阶高维线性系统. 因此，以下我们主要讨论一阶 m 维线性系统的结构性质.

在(2.1.3)式两端同乘一满秩方阵 P，有

$$\begin{aligned} P\boldsymbol{x}_t &= PA\boldsymbol{x}_{t-1} + P\boldsymbol{a} \\ &= PAP^{-1}P\boldsymbol{x}_{t-1} + P\boldsymbol{a}. \end{aligned}$$

记 $\tilde{\boldsymbol{x}}_t = P\boldsymbol{x}_t$，$\tilde{\boldsymbol{a}} = P\boldsymbol{a}$，$\tilde{A} = PAP^{-1}$，则上式为

$$\begin{cases} \tilde{\boldsymbol{x}}_t = \tilde{A}\tilde{\boldsymbol{x}}_{t-1} + \tilde{\boldsymbol{a}}; \\ \tilde{\boldsymbol{x}}_0 \in R^m. \end{cases} \tag{2.1.7}$$

根据矩阵知识，存在一个满秩方阵 P，使方阵 $\tilde{A}$ 具有如下约当标准形式：

$$\widetilde{A}=\begin{pmatrix} J_1 & & & 0 \\ & J_2 & & \\ & & \ddots & \\ 0 & & & J_s \end{pmatrix},$$

其中

$$J_k=\begin{pmatrix} \lambda_k & 1 & & 0 \\ & \lambda_k & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_k \end{pmatrix},$$

将 $\widetilde{\boldsymbol{a}}$ 按照 $\widetilde{A}$ 的对角块 $J_1,\cdots,J_s$ 的阶数分成如下分块向量：

$$\widetilde{\boldsymbol{a}}=(\widetilde{\boldsymbol{a}}_1^{\tau},\widetilde{\boldsymbol{a}}_2^{\tau},\cdots,\widetilde{\boldsymbol{a}}_s^{\tau})^{\tau},$$

其中 $\widetilde{\boldsymbol{a}}_k$ 的维数与 J_k 的阶数相同. 类似地，也将 $\widetilde{\boldsymbol{x}}_t$ 写成相应的分块向量：

$$\widetilde{\boldsymbol{x}}_t=(\widetilde{\boldsymbol{x}}_{1t}^{\tau},\widetilde{\boldsymbol{x}}_{2t}^{\tau},\cdots,\widetilde{\boldsymbol{x}}_{st}^{\tau})^{\tau}.$$

于是(2.1.7)式可写为

$$\begin{pmatrix} \widetilde{\boldsymbol{x}}_{1t} \\ \widetilde{\boldsymbol{x}}_{2t} \\ \vdots \\ \widetilde{\boldsymbol{x}}_{st} \end{pmatrix}=\begin{pmatrix} J_1 & & & 0 \\ & J_2 & & \\ & & \ddots & \\ 0 & & & J_s \end{pmatrix}\begin{pmatrix} \widetilde{\boldsymbol{x}}_{1,t-1} \\ \widetilde{\boldsymbol{x}}_{2,t-1} \\ \vdots \\ \widetilde{\boldsymbol{x}}_{s,t-1} \end{pmatrix}+\begin{pmatrix} \widetilde{\boldsymbol{a}}_1 \\ \widetilde{\boldsymbol{a}}_2 \\ \vdots \\ \widetilde{\boldsymbol{a}}_s \end{pmatrix}. \tag{2.1.8}$$

这样，(2.1.8)式成为 s 个相互分离的一阶线性系统，即

$$\widetilde{\boldsymbol{x}}_{kt}=J_k\widetilde{\boldsymbol{x}}_{k,t-1}+\widetilde{\boldsymbol{a}}_k \quad (k=1,2,\cdots,s).$$

对于不同的 k 值，J_k 的阶数可能不同. 但它们都具有如下形式：

$$\begin{cases} \boldsymbol{x}_t=J\boldsymbol{x}_{t-1}+\boldsymbol{a}; \\ \boldsymbol{x}_0\in R^j, \end{cases} \tag{2.1.9}$$

其中 J 为 j 阶约当矩阵：

$$J=\begin{pmatrix} \lambda & 1 & & 0 \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{pmatrix},$$

由此可见，为了讨论一阶高维线性系统(2.1.3)的结构性质，只需讨论线性系统(2.1.9)的结构性质即可. 当 $j=1$ 时，(2.1.8)式即为(2.1.1)式. 而当 $j>2$ 与 $j=2$ 时，在讨论中无实质性差别. 因此，我们仅讨论 $j=2$ 的

情况. 当 $j=2$ 时,(2.1.9)式的解的形式为:

$$\boldsymbol{x}_t = \boldsymbol{a} + J\boldsymbol{a} + \cdots + J^{t-1}\boldsymbol{a} + J^t\boldsymbol{x}_0, \qquad (2.1.10)$$

其中

$$J = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}.$$

注意到

$$J^k = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}^k = \begin{pmatrix} \lambda^k & k\lambda^{k-1} \\ 0 & \lambda^k \end{pmatrix} \quad (k = 1,2,\cdots),$$

则相应于(2.1.1)式的讨论不难推广到(2.1.10)式. 用 I 表示单位矩阵,其阶数与 J 的阶数相同. 容易验证:

(i) 当 $|\lambda|<1$ 时,对任何 $\boldsymbol{x}_0\in R^2$,有

$$\lim_{t\to\infty}\boldsymbol{x}_t=(I-J)^{-1}\boldsymbol{a};$$

(ii) 当 $|\lambda|>1$ 时,对任何 $\boldsymbol{x}_0\in R^2,\boldsymbol{x}_0\neq(I-J)^{-1}\boldsymbol{a}$,有

$$\lim_{t\to\infty}\|\boldsymbol{x}_t\|=\infty,$$

式中 $\|\cdot\|$ 表示向量的平方范数,即对向量 $\boldsymbol{u}=(u_1,\cdots,u_p)^{\tau}$,

$$\|\boldsymbol{u}\| = (|u_1|^2 + |u_2|^2 + \cdots + |u_p|^2)^{1/2};$$

当 $\boldsymbol{x}_0=(I-J)^{-1}\boldsymbol{a}$ 时,有

$$\boldsymbol{x}_t = \boldsymbol{x}_0 = (I - J)^{-1}\boldsymbol{a}.$$

对于性质 2.1.2 中的(iii)至(v)的推广,由于特征值 λ 可以是复数,向量 $\boldsymbol{a}$ 有两个分量,所以其情况比(2.1.1)式复杂些,可能不止三种情况. 但是,分析的方法是初等的,我们把它们留给读者考虑.

从以上讨论可见,有穷阶有限维的线性系统可转化为一阶高维线性系统. 而一阶高维线性系统的结构特征首先被它的参数矩阵的特征值和约当型所决定. 但这又由每个约当块确定的特殊线性系统(2.1.9)的结构所描述. 如上所述,由线性系统(2.1.9)迭代产生的序列,其极限特征与最简单的一阶一维线性系统(2.1.1)相比,其复杂程度并不太大. 而离散动态系统的结构特征主要是由迭代序列的各种极限性质所决定的. 例如,在本章§2.3 中将要定义的离散动态系统的不同的稳定性概念,就是根据不同的收敛性质决定的. 然而,将上述线性系统的讨论推广到非线性系统,即使是一阶一维的情况,不仅上述的分析方法不能奏效,而且其迭代

序列的极限情况也复杂得多，甚至于彻底解决一个简单的一阶一维非线性系统的极限分类问题，其难度都是无法估计的. 为此，我们将在下一节中简单地介绍一般离散动态系统的极限分类，然后在§2.3中讨论系统的稳定性概念.

§2.2　动态系统的吸引元和吸引域

在本节中，我们根据离散动态系统迭代序列的极限性质，定义两个描述离散动态系统结构特征的重要概念——吸引元和吸引域. 利用它们还可以引入§2.3中的稳定性概念，以进一步描述离散动态系统.

首先，我们给出§2.1中的线性系统(2.1.1)、(2.1.3)和(2.1.4)在非线性情况下的推广. 这里不再先从推广(2.1.1)做起，而是直接给出(2.1.3)式的推广，然后再给出(2.1.4)式的推广，而视(2.1.1)式为上述推广的特例.

定义 2.2.1　设 $\boldsymbol{T}(\boldsymbol{x})=(T_1(\boldsymbol{x}),T_2(\boldsymbol{x}),\cdots,T_m(\boldsymbol{x}))^{\tau}$ 为 R^m 到 R^m 的可测变换. 以下的迭代关系称为一阶 *m* 维动态系统：

$$\begin{cases}\boldsymbol{x}_t=\boldsymbol{T}(\boldsymbol{x}_{t-1}) \quad (t\geqslant 1);\\ \boldsymbol{x}_0\in R^m.\end{cases} \tag{2.2.1}$$

在上式中，如果 m 维变换 $\boldsymbol{T}(\boldsymbol{x})$ 是 R^m 中的一个子集 D 到其自身的可测变换，则(2.2.1)式应写为：

$$\begin{cases}\boldsymbol{x}_t=\boldsymbol{T}(\boldsymbol{x}_{t-1}) \quad (t\geqslant 1);\\ \boldsymbol{x}_0\in D,\end{cases} \tag{2.2.2}$$

其中 $D\subset R^m$. 显然，当 $D=R^m$ 时，系统(2.2.2)即为(2.2.1). 因此，以下我们将不加区别地使用系统(2.2.1)和(2.2.2)，视其讨论问题的方便而定. 为了不涉及更多的数学领域，这里只考虑 D 为 R^m 中非降维的凸子集.

当 $\boldsymbol{T}(\boldsymbol{x})=A\boldsymbol{x}+\boldsymbol{a}$ 时，(2.2.1)式即为线性系统(2.1.3). 当 A 还具有(2.1.5)式的形式，且 $\boldsymbol{a}=(a_0,0,\cdots,0)^{\tau}$ 时，(2.2.1)式即为线性系统(2.1.6).

定义 2.2.2　设 φ 是 R^p 到 R^1 的可测变换，或者说是从 R^p 到 R^1 的 p 元标量函数. 以下的迭代关系称为 *p* 阶一维动态系统：

$$\begin{cases}x_t=\varphi(x_{t-1},\cdots,x_{t-p}) \quad (t\geqslant 1);\\ (x_0,x_{-1},\cdots,x_{-p+1})\in R^p.\end{cases} \tag{2.2.3}$$

特别地，当变换 φ 为线性函数时，即

$$\varphi(x_{t-1},\cdots,x_{t-p}) = a_0 + a_1x_{t-1} + \cdots + a_px_{t-p},$$

(2.2.3)式即为 p 阶一维线性系统(2.1.4).

为了将(2.2.3)式写成(2.2.1)式的等价形式,我们令

$$\boldsymbol{x}_t = (x_t, x_{t-1}, \cdots, x_{t-p+1})^\tau,$$

$$\boldsymbol{T}(\boldsymbol{x}_t) = (\varphi(\boldsymbol{x}_t), x_t, x_{t-1}, \cdots, x_{t-p+2})^\tau.$$

于是(2.2.3)式可写为

$$\begin{cases} \boldsymbol{x}_t = \boldsymbol{T}(\boldsymbol{x}_{t-1}) \quad (t \geqslant 1); \\ \boldsymbol{x}_0 \in R^p. \end{cases} \tag{2.2.4}$$

如同前一节所述,任何 p 阶 m 维线性系统均可转化为一阶 pm 维线性系统.对于非线性系统可进行类似的转化.所以从理论上说,研究一阶多维动态系统具有一定的普遍意义.

在§2.1中,线性系统(2.1.3)可以用矩阵运算进行等价变换.在本节中,对非线性系统(2.2.1)不能再用同样方法进行变换.类似地,我们可以考虑如下变换:设 P 是从 D 到 $\widetilde{D}\subset R^m$ 的一对一的可逆变换.记 $\tilde{\boldsymbol{x}}=P(\boldsymbol{x})$, $\widetilde{\boldsymbol{T}}=PTP^{-1}$为复合变换.于是,(2.2.2)式等价于

$$\begin{cases} \tilde{\boldsymbol{x}}_t = \widetilde{\boldsymbol{T}}(\tilde{\boldsymbol{x}}_{t-1}) \quad (t \geqslant 1); \\ \tilde{\boldsymbol{x}}_0 \in \widetilde{D}. \end{cases} \tag{2.2.5}$$

但是,我们很难找到一个适当的变换 P,使得(2.2.5)式比(2.2.2)式更简单,甚至不可能有任何实质上的简化.不过上述变换可将无界区域 D 转化成有界区域 $\widetilde{D}$,使得讨论(2.2.5)式中的变换 $\widetilde{\boldsymbol{T}}$ 的结构特征能被限制在有界区域上,例如单位圆到单位圆上.

在§2.1中,不仅可将(2.1.3)式转化成(2.1.7)式,而且对(2.1.3)式结构特征的讨论也可转化为仅需讨论参数矩阵 A 具有约当型的(2.1.9)式,并有显式解的表达式(2.1.10).但对非线性系统(2.2.3),则既不能使等价变换形式(2.2.5)式更简化,也没有类似于(2.1.10)式的解的表达式.由此可看到线性系统与非线性系统的差异之大了.

为了以下讨论方便,我们引入形式上的迭代序列记号如下:

$$\begin{aligned}
&\boldsymbol{x}_0 = \boldsymbol{T}_0(\boldsymbol{x}_0), \\
&\boldsymbol{x}_1 = \boldsymbol{T}(\boldsymbol{x}_0) = \boldsymbol{T}_1(\boldsymbol{x}_0), \\
&\boldsymbol{x}_2 = \boldsymbol{T}(\boldsymbol{x}_1) = \boldsymbol{T}(\boldsymbol{T}(\boldsymbol{x}_0)) = \boldsymbol{T}_2(\boldsymbol{x}_0), \\
&\cdots\cdots\cdots \\
&\boldsymbol{x}_t = \boldsymbol{T}(\boldsymbol{x}_{t-1}) = \boldsymbol{T}_2(\boldsymbol{x}_{t-2}) = \cdots = \boldsymbol{T}_t(\boldsymbol{x}_0),
\end{aligned}$$

上式中 $\boldsymbol{T}_t(\boldsymbol{x}_0)$ 表示从 $\boldsymbol{x}_0$ 出发，用变换 $\boldsymbol{T}$ 迭代 t 次后的值.

刻划动态系统(2.2.1)的结构特征的方法不止一种. 将(2.2.1)式视为差分方程，如果能够给出(2.2.1)式的解的表达式，则将有助于分析动态系统的结构特征. 如 § 2.1 中的线性情况即是如此. 在非线性情况下，除了特殊变换 $\boldsymbol{T}$ 以外，非线性差分方程(2.2.1)式都没有显式解的表达式. 动态系统(2.2.1)的结构特征主要是通过迭代序列 $\{\boldsymbol{T}_t(\boldsymbol{x}_0)\}$ 表现出来. 与第 1 章的自回归模型(1.3.6)式相比较，那里的迭代序列是一个随机过程的样本序列，也称为*样本轨道*. 动态系统(2.2.1)是一个确定性的差分方程，它的一个迭代序列 $\{\boldsymbol{x}_t=\boldsymbol{T}_t(\boldsymbol{x}_0), t\geqslant 0\}$ 可称为动态系统(2.2.1)的*一条轨道*或*一次实现*. 此时，它是由变换 $\boldsymbol{T}$ 和初始值 $\boldsymbol{x}_0$ 唯一决定的确定性序列. 用这些不同轨道的极限性质描述该动态系统的结构是刻划该动态系统的第一步，也是很重要的一步. 粗略地说，首先要理清楚动态系统把不同的初始状态 $\boldsymbol{x}_0$ 最终运动到何处. 对于线性系统，按照参数所属的区域，从不同的初始值出发，便可确定出轨道的极限特征. 在非线性情况下，非线性动态系统的轨道所表现的极限行为比线性系统复杂得多. 在本节中，我们要分类介绍不同的极限行为，它们是描述非线性动态系统结构特征的重要方面.

定义 2.2.3　对动态系统(2.2.1)而言：

(i) 若由(2.2.1)式产生的迭代序列 $\{\boldsymbol{T}_t(\boldsymbol{x}_0), t\geqslant 0\}$ 满足

$$\lim_{t\to\infty}\boldsymbol{T}_t(\boldsymbol{x}_0)=\boldsymbol{a}, \tag{2.2.6}$$

则称 $\boldsymbol{a}$ 为动态系统(2.2.1)的*极限点*，也称为变换 $\boldsymbol{T}$ 的极限点；

(ii) 若存在 q 个不同的向量 $\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_q$，使得由(2.2.1)式产生的迭代序列 $\{\boldsymbol{T}_t(\boldsymbol{x}_0), t\geqslant 0\}$ 满足

$$\lim_{k\to\infty}\boldsymbol{T}_{kq+j}(\boldsymbol{x}_0)=\boldsymbol{a}_j \quad (j=1,2,\cdots,q), \tag{2.2.7}$$

则称 $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_q\}$ 为动态系统(2.2.1)的*极限环*，也称为变换 $\boldsymbol{T}$ 的极限环. q 称为*周期长度*. 当 $q=1$ 时，极限环 $\{\boldsymbol{a}_1\}$ 即为极限点.

动态系统(2.2.1)的迭代序列 $\{\boldsymbol{T}_t(\boldsymbol{x}_0), t\geqslant 0\}$ 的所有可能的极限值的全体，记为 $L(\boldsymbol{x}_0,\boldsymbol{T})$，即

$$L(\boldsymbol{x}_0,\boldsymbol{T})=\{\boldsymbol{x}:\text{存在子序列 }\boldsymbol{T}_{n_k}(\boldsymbol{x}_0)\text{，使得 }\boldsymbol{T}_{n_k}(\boldsymbol{x}_0)\to\boldsymbol{x}, n_k\to\infty\}. \tag{2.2.8}$$

根据定义 2.2.3 可知，如果 $\boldsymbol{a}$ 是变换 $\boldsymbol{T}$ 的极限点，则存在 $\boldsymbol{x}_0$，使得 $L(\boldsymbol{x}_0,\boldsymbol{T})=\{\boldsymbol{a}\}$. 若 $\{\boldsymbol{a}_1,\cdots,\boldsymbol{a}_q\}$ 是变换 $\boldsymbol{T}$ 的极限环，则存在 $\boldsymbol{x}_0$，使 $L(\boldsymbol{x}_0,\boldsymbol{T})=$

$\{\boldsymbol{a}_1,\cdots,\boldsymbol{a}_q\}$. 但是 $L(\boldsymbol{x}_0,\boldsymbol{T})$ 可能含有无穷多个元素. 为此,我们引入如下概念:

定义 2.2.4 对于动态系统(2.2.1)而言:

(i) 如果集合 $B=L(\boldsymbol{x}_0,\boldsymbol{T})$,则称 B 为该动态系统的吸引元,也称为变换 $\boldsymbol{T}$ 的吸引元;

(ii) 如果吸引元 B 为 Lebesgue 零测集,则称 B 为奇异吸引元;

(iii) 如果吸引元 B 的任何真子集都不再是吸引元,则称 B 为单纯吸引元.

定义 2.2.5 对动态系统(2.2.1),记

$$L(\boldsymbol{T}) = \bigcup_{\boldsymbol{x}_0\in R^m} L(\boldsymbol{x}_0,\boldsymbol{T}), \tag{2.2.9}$$

则 $L(\boldsymbol{T})$ 称为该动态系统的吸引集.

定义 2.2.6 对于动态系统(2.2.1)中的 $\boldsymbol{T}$ 而言:

(i) 如果 $\boldsymbol{a}\in R^m$ 满足 $\boldsymbol{T}(\boldsymbol{a})=\boldsymbol{a}$,则称 $\boldsymbol{a}$ 为变换 $\boldsymbol{T}$ 的不动点;

(ii) 如果存在 q 个不同的向量 $\boldsymbol{a}_1,\cdots,\boldsymbol{a}_q$,使得

$$\boldsymbol{a}_{j+1} = \boldsymbol{T}(\boldsymbol{a}_j) \quad (j = 1,\cdots,q),\text{且 } \boldsymbol{a}_{q+1} = \boldsymbol{a}_1, \tag{2.2.10}$$

则称 $\{\boldsymbol{a}_1,\cdots,\boldsymbol{a}_q\}$ 为变换 $\boldsymbol{T}$ 的周期解.

关于吸引元的如下几条性质,其结论是显然的,我们略去证明.

性质 2.2.7 对动态系统(2.2.1),有

(i) 极限点和极限环都是动态系统的奇异吸引元.

(ii) 动态系统中变换 $\boldsymbol{T}$ 的不动点必为极限点. 反之,如果 $\boldsymbol{a}$ 是变换 $\boldsymbol{T}$ 的极限点,且变换 $\boldsymbol{T}$ 在 $\boldsymbol{a}$ 处连续,则 $\boldsymbol{a}$ 必是 $\boldsymbol{T}$ 的不动点.

(iii) 动态系统中变换 $\boldsymbol{T}$ 的周期解必为 $\boldsymbol{T}$ 的极限环. 反之,如果 $\{\boldsymbol{a}_1,\cdots,\boldsymbol{a}_q\}$ 是变换 $\boldsymbol{T}$ 的极限环,且 $\boldsymbol{T}$ 在 $\boldsymbol{a}_1,\cdots,\boldsymbol{a}_q$ 处连续,则 $\{\boldsymbol{a}_1,\cdots,\boldsymbol{a}_q\}$ 必为变换 $\boldsymbol{T}$ 的周期解.

当吸引元是奇异吸引元,但又不是极限点或极限环时,又有两种不同情况. 一是此吸引元属于 R^m 中的某个低维流型的子集;另一是此吸引元不属于 R^m 中的任何低维流型. 对于极限点和极限环以外的吸引元的分类讨论,有着丰富的内容. 本章作为全书的预备知识,仅限于论及以上关于吸引元的较初等的分类. 在第 8 章中,为了阐明非线性时间序列与相关学科的联系,还要简要地论及吸引元的更复杂的性质的描述.

下面举例说明动态系统吸引元的概念. 首先分析最简单的线性系统(2.1.1). 易见,在性质 2.1.2 中(i)和(ii)的情况下,$a_0/(1-a_1)$ 是该线性

系统的唯一吸引元，且为极限点. 在性质 2.1.2 中(iv)的情况下，每个实数都是系统的极限点，且是不动点. 在性质 2.1.2 中(v)的情况下，对任何实数 $x\neq a_0/2$，$\{-x,x\}$都是系统的周期长度为 2 的极限环. 在性质2.1.2 中(iii)的情况下，由于没有极限，可以说无穷远点是一个极限点. 类似地，对一般线性系统，按照 § 2.1 的论述，也不难分析清楚其极限点和极限环出现的条件. 对线性系统而言，它只有极限点和极限环，而没有其他类型的吸引元. 从下面的例子可以看出，这也是线性系统与非线性系统的差别之一.

例 2.2.8　考虑如下的一维分段线性系统：

$$x_t=\begin{cases}\beta x_{t-1}, & \text{当 } x_{t-1}\leqslant 1;\\ \alpha x_{t-1}, & \text{当 } x_{t-1}>1,\end{cases}\tag{2.2.11}$$

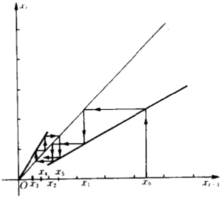


图 2.2.1(a)　动态系统(2.2.11)的迭代示意图

其中 $\beta=\frac{3}{2}$，$\alpha=\frac{1}{2}$

其中参数 $\beta>1$，$0<\alpha<1$，$x_0\in(0,\infty)$，当 $\log\alpha/\log\beta$（这里 log 表示取自然对数）为无理数时，由 § 8.2可知，对任何 $x_0>0$，迭代序列$\{x_t,t\geqslant 0\}$为区间$[\alpha,\beta]$中的稠密子集. 于是由(2.2.8)式和定义2.2.4知$[\alpha,\beta]$是系统(2.2.11)的一个吸引元. 又由于 $x_0\in(0,\infty)$是任意的，故$[\alpha,\beta]$是系统(2.2.11)的唯一的吸引元，从而是单纯吸引元.

如果将(2.2.11)的两端取对数，并令 $\tilde{x}_t=\log x_t$，$\tilde{\alpha}=\log\alpha$，$\tilde{\beta}=\log\beta$，于是得到如下的非线性系统：

$$\tilde{x}_t=\begin{cases}\tilde{x}_{t-1}+\tilde{\beta}, & \text{当 } \tilde{x}_{t-1}\leqslant 0;\\ \tilde{x}_{t-1}+\tilde{\alpha}, & \text{当 } \tilde{x}_{t-1}>0,\end{cases}\tag{2.2.12}$$

其中 $\tilde{\beta}>0$. $\tilde{\alpha}<0$. 根据例 2.2.8 的讨论可知，当 $\tilde{\alpha}/\tilde{\beta}$ 为无理数时，系统(2.2.12)有唯一的单纯吸引元$[\tilde{\alpha},\tilde{\beta}]$.

例 2.2.9　考虑如下的分段线性系统：

$$\begin{cases}x_t=2x_{t-1}\bmod(1) \quad (t\geqslant 1);\\ x_0\in[0,1],\end{cases}\tag{2.2.13}$$

其中变换

$$T(x) = 2x\bmod(1) = \begin{cases} 2x, & \text{当 } 0 \leqslant x \leqslant \dfrac{1}{2}; \\ 2x - 1, & \text{当 } \dfrac{1}{2} < x \leqslant 1. \end{cases} \quad (2.2.14)$$

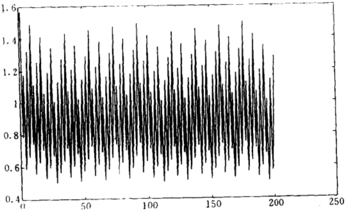


图 2.2.1(b) 动态系统(2.2.11)的迭代序列轨道

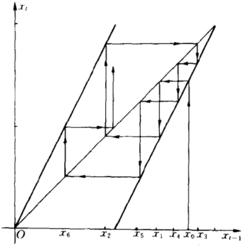


图 2.2.2(a) 动态系统(2.2.13)的迭代示意图

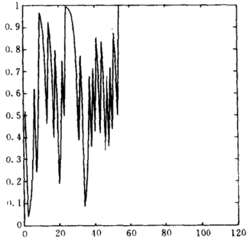


图 2.2.2(b) 动态系统(2.2.13)的迭代序列轨道

由于 $T(0)=0$,$T(1)=1$,而且 $T(x)$ 在 $x=0$ 和 1 处为连续的,根据性质 2.2.7 知,{0}和{1}是系统(2.2.13)的两个不动点. 又易见,$\{1/3, 2/3\}$ 为 $T(x)$ 的周期解,而且 $T(x)$ 在 $x=1/3$ 和 $2/3$ 处为连续的,故 $\{1/3, 2/3\}$ 是系统(2.2.13)的极限环. 实际上,从任何有理数 x_0 出发,迭代序列 $\{x_t=$

$T_t(x_0), t \geqslant 0\}$的尾部都是周期序列，即从某个正整数 n 以后，$\{x_t, t \geqslant n\}$是一个周期序列. 由此可见，系统(2.2.13)有可数多个极限环。另外，区间$[0,1]$也是系统(2.2.13)的吸引元. 不过，讨论这一内容要涉及混沌概念，我们将在第 8 章叙述. 这里只在于说明，吸引元$[0,1]$不是单纯吸引元，因为它包含了其他吸引元作为其子集.

例 2.2.10　考虑如下的非线性系统：

$$x_t = \begin{cases} \frac{1}{2} x_{t-1}, & \text{当 } x_{t-1} \neq \text{正奇数}; \\ 3x_{t-1} + 1, & \text{当 } x_{t-1} = \text{正奇数}, \end{cases} \tag{2.2.15}$$

$$x_0 \in [0, \infty).$$

易见，$\{0\}$是系统(2.2.15)的一个极限点. $\{1,4,2\}$是系统(2.2.15)的一个极限环. 试问，此系统除了上述两个吸引元外，还有没有其他吸引元？又问，系统(2.2.15)是否有趋于无穷的迭代序列？根据 50 多年前的一个猜想，系统(2.2.15)是没有其他吸引元的，也没有发散的迭代序列. 这一猜想是对如下正整数迭代系统提出的，即

$$n_t = \begin{cases} \frac{1}{2} n_{t-1}, & \text{当 } n_{t-1} \text{ 为偶数}; \\ 3n_{t-1} + 1, & \text{当 } n_{t-1} \text{ 为奇数}, \end{cases} \tag{2.2.16}$$

其中初始值 n_0 为任意正整数. 人们猜想(见文献[72])，无论 n_0 取怎样的正整数，由(2.2.16)式产生的正整数序列，其尾部都是由$\{1,4,2\}$三个元素组成的周期序列. 这一猜想被许多学者研究过，但至今没有任何实质性的进展。尽管有人用计算机验证了在 $1 \leqslant n_0 \leqslant 2^{50}$ 内猜想是成立的(见文献[72]). 但是，既没有证明迭代序列无发散现象，也没有证明系统(2.2.16)只有一个极限环. 上述猜想被称为 $3x+1$ 问题(见文献[72]). 50 多年来，此问题未能被解决，足以说明其难度之大. 更有甚者，曾有专家认为“解决这样问题的数学还没有问世”(见文献[72]). 在此举出此例，仅仅是借以说明研究非线性系统的困难程度实在是无法估量的.

例 2.2.11　考虑如下的动态系统：

$$\begin{cases} x_t = \dfrac{\alpha x_{t-1}}{1 + x_{t-1}^2}; \\ x_0 \in [0, \infty). \end{cases} \tag{2.2.17}$$

此动态系统中的变换 $T(x) = \alpha x/(1+x^2)$的图形见图 2.2.3. 此动态系统的吸引元与 α 的取值有关，下面分不同情况分别加以讨论. 由于 $\alpha < 0$ 的

情况与 $\alpha>0$ 的情况类似，以下只讨论 $\alpha>0$ 的情况.

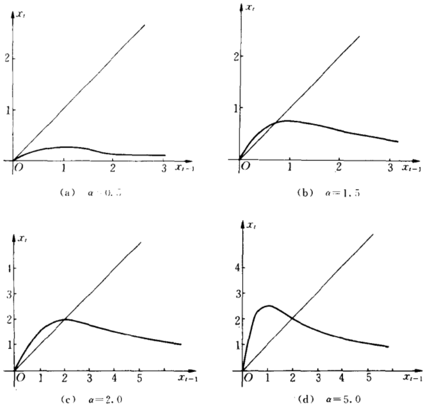


图 2.2.3　动态系统(2.2.17)中变换的图形

(i) 当 $0<\alpha\leqslant 1$ 时，动态系统(2.2.17)的极限点即为 $T(x)=x$ 的不动点，只有$\{0\}$. 设 $x_0=x\in[0,\infty)$，于是由(2.2.17)式知

$$0<x_1=\frac{\alpha x_0}{1+x_0^2}<x_0,$$

从而又有

$$0<x_2=\frac{\alpha x_1}{1+x_1^2}<x_1.$$

进而对任意 t，有

$$0<x_t=\frac{\alpha x_{t-1}}{1+x_{t-1}^2}<x_{t-1}\quad(t=1,2,\cdots).$$

由于 $T(x)=x$ 只有一个不动点$\{0\}$，所以，对任何 $x_0=x\in[0,\infty)$，迭代序列都趋于 0. 这说明动态系统(2.2.17)的吸引元只有不动点$\{0\}$.

(ii) 当 $1<\alpha<2$ 时，利用对 $T(x)$的导数

$$T'(x) = \alpha(1 - x^2)/(1 + x^2)^2 \tag{2.2.18}$$

的分析可知，$T(x)$ 在 $x=1$ 处取极大值 $T(1)=\alpha/2$. 这对任意 α 均成立. 由图 2.2.3 可见，$T(x)=x$ 有两个解，即 $\{0\}$ 和 $\{(\alpha-1)^{1/2}\}$，它们都是动态系统(2.2.17)的极限点.

若 $x_0>1$，则由(2.2.17)式知

$$x_1 = \frac{\alpha x_0}{1 + x_0^2} \leqslant \frac{\alpha}{2} < 1,$$

从而，不妨只讨论 $x_0 \leqslant 1$ 的情况. 若 $0<x_0<(\alpha-1)^{1/2}$，由(2.2.17)式知

$$x_1 = \frac{\alpha x_0}{1 + x_0^2} > \frac{\alpha x_0}{1 + (\alpha - 1)} = x_0.$$

在此注意，$T(x)$ 在 $(0,(\alpha-1)^{1/2})$ 中为单调上升的，从而又有

$$x_1 = T(x_0) < T((\alpha - 1)^{1/2}) = (\alpha - 1)^{1/2}.$$

以上两式说明了 $0<x_1<(\alpha-1)^{1/2}$，且 $x_1>x_0$. 由类似分析有

$$0 < x_2 < (\alpha - 1)^{1/2}, \quad x_2 > x_1.$$

依次类推，有

$$0 < x_t < (\alpha - 1)^{1/2}, \quad x_t > x_{t-1}.$$

上式表明了 $\{x_t\}$ 为单调上升序列，且都小于 $(\alpha-1)^{1/2}$，从而必有极限. 又由于 $T(x)$ 除了 $\{0\}$ 以外，只有不动点 $\{(\alpha-1)^{1/2}\}$，因而 $\{x_t\}$ 必收敛于 $(\alpha-1)^{1/2}$.

若 $(\alpha-1)^{1/2} \leqslant x_0 \leqslant 1$，依上述类似的方法得知，由(2.2.17)式迭代的序列必然下降趋于 $(\alpha-1)^{1/2}$.

综合以上所述，在 $1<\alpha<2$ 时，动态系统(2.2.17)只有吸引元 $\{0\}$ 和 $\{(\alpha-1)^{1/2}\}$，它们是两个不同的极限点.

(iii) 当 $\alpha=2$ 时，由图 2.2.3 可见，$T(x)=2x/(1+x^2)$ 在 $x=1$ 处既为极值点，又为不动点. 用(ii)中类似的论述方法可知，此时动态系统(2.2.17)只有两个极限点 $\{0\}$ 和 $\{1\}$.

(iv) 当 $\alpha>2$ 时，$T(x)=\alpha x/(1+x^2)$ 的极大值点仍在 $x=1$ 处，其不动点为 $\{0\}$ 和 $\{(\alpha-1)^{1/2}\}$，此时 $(\alpha-1)^{1/2}>1$.

若 $0<x_0<1$，则由(2.2.17)式知

$$x_1 = \frac{\alpha x_0}{1 + x_0^2} > \frac{\alpha x_0}{1 + 1} = \frac{\alpha}{2} x_0 > x_0.$$

如果 $x_1<1$，由上式类似结果有 $x_2>x_1$. 此事实表明了从 $0<x_0<1$ 出发，由(2.2.17)式迭代得到 $x_0, x_1, \cdots$，当 x_t 在首次超过 1 之前，此序列是单调上升的. 由于在 $(0,1)$ 之间 $T(x)$ 没有不动点，所以序列 $\{x_t\}$ 必有某个 x_t

达到或超过 1 值. 依此,我们不妨只讨论 $x_0 \geqslant 1$ 的情况.

若 $x_0 \geqslant 1$,则由于 $T(x_0) \leqslant T(1) = \alpha/2$,从而序列 $\{x_t\}$ 满足

$$x_t = T(x_{t-1}) \leqslant T(1) = \alpha/2, \quad t \geqslant 1. \tag{2.2.19}$$

此式告诉我们,只需考虑 $1 \leqslant x_0 \leqslant \alpha/2$ 的情况.

若 $1 \leqslant x_0 \leqslant \alpha/2$,由(2.2.19)式知,对一切 x_t,(2.2.19)式成立. 再由(2.2.18)式可知,$T(x)$ 在 $(1, \alpha)$ 上为单调下降的,从而由 $1 \leqslant x_0 \leqslant \alpha/2$ 和 $\alpha > 2$ 可得到

$$x_1 = T(x_0) \geqslant T(\alpha/2)$$

$$= \frac{\alpha \cdot \dfrac{\alpha}{2}}{1 + \dfrac{\alpha^2}{4}} = \frac{2\alpha^2}{4 + \alpha^2} > 1.$$

联合上式和(2.2.19)式知 $1 \leqslant x_1 \leqslant \alpha/2$. 重复以上推论,可得

$$1 \leqslant x_t \leqslant \alpha/2, \quad t \geqslant 0.$$

利用 $T(x)$ 在 $[1, \alpha/2]$ 中单调下降以及 $T((\alpha-1)^{1/2}) = (\alpha-1)^{1/2}$,易见

$$x_t \begin{cases} \leqslant (\alpha - 1)^{1/2}, & \text{当 } x_{t-1} \geqslant (\alpha - 1)^{1/2}; \\ \geqslant (\alpha - 1)^{1/2}, & \text{当 } x_{t-1} \leqslant (\alpha - 1)^{1/2}. \end{cases} \tag{2.2.20}$$

上式表明

$$x_{2t} \begin{cases} \leqslant (\alpha - 1)^{1/2}, & \text{当 } x_{2t-2} \leqslant (\alpha - 1)^{1/2}; \\ \geqslant (\alpha - 1)^{1/2}, & \text{当 } x_{2t-2} \geqslant (\alpha - 1)^{1/2}. \end{cases} \tag{2.2.21}$$

现在考虑 $\{x_t\}$ 的子序列 $\{x_{2t}\}$,即

$$\begin{aligned} x_{2t} &= T(T(x_{2t-2})) \\ &= \frac{\alpha^2(1 + x_{2t-2}^2)}{(1 + x_{2t-2}^2)^2 + \alpha^2 x_{2t-2}^2} \cdot x_{2t-2} \\ &= \psi(x_{2t-2}) x_{2t-2}, \end{aligned} \tag{2.2.22}$$

其中

$$\psi(x) = \frac{\alpha^2(1 + x^2)}{(1 + x^2)^2 + \alpha^2 x^2}, \quad 1 \leqslant x \leqslant \alpha - 1.$$

易见,$\psi(x)$ 在 $x \in [0, \infty)$ 上为单调下降的. 这样,若 $x_0 \leqslant (\alpha-1)^{1/2}$,由以上两式,有

$$\begin{aligned} (\alpha - 1)^{1/2} \geqslant x_2 = \psi(x_0) x_0 &\geqslant \psi((\alpha - 1)^{1/2}) \cdot x_0 \\ &= x_0. \end{aligned}$$

继续以上推理可知

$$(\alpha-1)^{1/2} \geqslant x_{2t} = \psi(x_{2t-2})x_{2t-2} \geqslant \psi((\alpha-1)^{1/2}) \cdot x_{2t-2}$$
$$= x_{2t-2} \quad (t \geqslant 1). \tag{2.2.23}$$

根据 $T(x)=x$ 在 $x\geqslant 1$ 时只有一个解 $(\alpha-1)^{1/2}$，由上式可知，$\{x_{2t}\}$ 必趋于 $(\alpha-1)^{1/2}$.

利用(2.2.20)式和(2.2.21)式不难证明，$\{x_{2t+1}\}$ 也必趋于 $(\alpha-1)^{1/2}$. 再用完全类似的方法可以证明，若 $1<x_0\leqslant(\alpha-1)^{1/2}$，$\{x_t\}$ 仍趋于 $(\alpha-1)^{1/2}$.

综上所述，在 $\alpha>2$ 时，动态系统(2.2.17)仍只有两个不动点 $\{0\}$ 和 $\{(\alpha-1)^{1/2}\}$ 为其吸引元.

例 2.2.12　考虑如下的二阶一维迭代系统：

$$\begin{cases} x_t = \left\{\dfrac{1}{2}(x_{t-1}^2 + x_{t-2}^2)\right\}^{1/2}; \\ (x_0, x_{-1}) \in D = [0,1] \times [0,1]. \end{cases} \tag{2.2.24}$$

记
$$\boldsymbol{x}_t = (x_t, x_{t-1})^\tau,$$
$$\varphi(x_1, x_2) = \left\{\frac{1}{2}(x_1^2 + x_2^2)\right\}^{1/2},$$
$$\boldsymbol{T}(x_1, x_2) = (\varphi(x_1, x_2), x_1)^\tau,$$

则(2.2.24)式可写成一阶二维动态系统：

$$\begin{cases} \boldsymbol{x}_t = \boldsymbol{T}(\boldsymbol{x}_{t-1}); \\ \boldsymbol{x}_0 \in D = [0,1] \times [0,1]. \end{cases} \tag{2.2.25}$$

易见，$\boldsymbol{T}$ 为从 D 到 D 上的连续变换. 又易见，当 $0<c<1$ 时，$(c,c)^\tau$ 为 $\boldsymbol{T}$ 的不动点，从而也是 $\boldsymbol{T}$ 的极限点.

现在考虑从 $x_0=b$，$x_{-1}=a$ 出发的迭代序列 $\{x_t\}$，不妨先假定 $0\leqslant a<b\leqslant 1$，于是有

$$\begin{cases} x_t^2 = \dfrac{1}{2}(x_{t-1}^2 + x_{t-2}^2); \\ x_0^2 = b^2, \ x_{-1}^2 = a^2. \end{cases}$$

上式是关于 x_t^2 的线性差分方程. 由直接验证可知，它有如下的解：

$$x_t^2 = \frac{2}{3}b^2 + \frac{1}{3}a^2 + \frac{1}{3}(b^2 - a^2)\left(-\frac{1}{2}\right)^t$$
$$(t = -1, 0, 1, \cdots).$$

由此可知，当 $x_0=b>x_{-1}=a$ 时，有

$$\lim_{t\to\infty} x_t = \left(\frac{2}{3}b^2 + \frac{1}{3}a^2\right)^{1/2}.$$

当 $x_0=b<x_{-1}=a$ 时，由(2.2.24)式有

$$x_1=\left\{\frac{1}{2}(b^2+a^2)\right\}^{1/2}>b=x_0.$$

于是由(2.2.24)式迭代所得的序列必有极限：

$$\begin{aligned}\lim_{t\to\infty}x_t&=\left(\frac{2}{3}x_1^2+\frac{1}{3}x_0^2\right)^{1/2}\\&=\left(\frac{2}{3}b^2+\frac{1}{3}a^2\right)^{1/2}.\end{aligned}$$

从而可知，从 $x_0=b$，$x_{-1}=a$ 出发，由(2.2.24)式产生的迭代序列 $\{x_t\}$ 总有

$$\lim_{t\to\infty}x_t=\left(\frac{2}{3}b^2+\frac{1}{3}a^2\right)^{1/2}. \tag{2.2.26}$$

换言之，记 $\boldsymbol{x}_0=(b,a)^{\tau}\in D$，则由(2.2.25)和(2.2.26)式，有

$$\lim_{t\to\infty}\boldsymbol{x}_t=\left(\left(\frac{2}{3}b^2+\frac{1}{3}a^2\right)^{1/2},\ \left(\frac{2}{3}b^2+\frac{1}{3}a^2\right)^{1/2}\right)^{\tau}.$$

上式表明了系统(2.2.25)的吸引集为

$$L=\{(a,a)^{\tau};0\leqslant a\leqslant 1\},$$

即 L 为由[0,1]×[0,1]的对角线元素组成，而且每个元素都是 $\boldsymbol{T}$ 的极限点. 由此可知，此吸引集 L 是 D 中的降维子集.

以上我们介绍并讨论了动态系统的吸引元的概念. 顾名思义，吸引元就是系统的某些轨道的极限归宿. 下面要讨论的是一个吸引元能够吸引哪些初始状态的问题.

定义 2.2.13 设 B 是动态系统(2.2.1)的一个吸引元，记

$$A_T(B)=\{\boldsymbol{x}:r_t(\boldsymbol{x},B)\to 0,\ t\to\infty\}, \tag{2.2.27}$$

其中 $r_t(\boldsymbol{x},B)$ 表示点 $\boldsymbol{T}_t(\boldsymbol{x})$ 到集合 B 的普通欧氏空间的距离，即

$$r_t(\boldsymbol{x},B)=r(\boldsymbol{T}_t(\boldsymbol{x}),B)=\inf_{y\in B}\|\boldsymbol{T}_t(\boldsymbol{x})-\boldsymbol{y}\|, \tag{2.2.28}$$

$A_T(B)$ 称为 B 的**吸引域**.

直观地说，对于吸引域 $A_T(B)$ 中的每个元素 $\boldsymbol{x}\in A_T(B)$，动态系统(2.2.1)以 $\boldsymbol{x}_0=\boldsymbol{x}$ 为初始值产生的迭代序列最终将进入或任意接近吸引元 B，就像被 B 吸引住一样.

性质 2.2.14 如果动态系统(2.2.1)中的变换 $\boldsymbol{T}$ 是连续的，则任何有界的吸引元 B 都是闭集，且

$$B\subset A_T(B). \tag{2.2.29}$$

证明 由于 B 是吸引元，依定义，存在 $\boldsymbol{x}_0\in R^m$，使得

$$B = L(\boldsymbol{x}_0, \boldsymbol{T}).$$

若 $\boldsymbol{b}_k \in B$，且 $\boldsymbol{b}_k \to \boldsymbol{b}$，则存在$\{\boldsymbol{T}_t(\boldsymbol{x}_0)\}$的子序列$\{\boldsymbol{T}_{n_k}(\boldsymbol{x}_0)\}$，使得

$$\| \boldsymbol{T}_{n_k}(\boldsymbol{x}_0) - \boldsymbol{b}_k \| \leqslant \frac{1}{k}.$$

于是

$$\| \boldsymbol{T}_{n_k}(\boldsymbol{x}_0) - \boldsymbol{b} \| \leqslant \| \boldsymbol{T}_{n_k}(\boldsymbol{x}_0) - \boldsymbol{b}_k \| + \| \boldsymbol{b}_k - \boldsymbol{b} \| \to 0.$$

此式表明 $\boldsymbol{b}$ 是$\{\boldsymbol{T}_{n_k}(\boldsymbol{x}_0)\}$的极限，从而 $\boldsymbol{b} \in B$，亦即 B 是闭集.

又对任意 $\boldsymbol{b} \in B$，由 $\boldsymbol{T}$ 的连续性及前一步的证明可知，存在 $\boldsymbol{x}_0 \in R^m$ 和$\{\boldsymbol{T}_t(\boldsymbol{x}_0)\}$的子序列$\{\boldsymbol{T}_{n_k}(\boldsymbol{x}_0)\}$，使得

$$\boldsymbol{T}_{n_{k+1}}(\boldsymbol{x}_0) = \boldsymbol{T}(\boldsymbol{T}_{n_k}(\boldsymbol{x}_0)) \to \boldsymbol{b}.$$

上式表明，$\boldsymbol{T}(\boldsymbol{b})$也是$\{\boldsymbol{T}_t(\boldsymbol{x}_0)\}$的子序列$\{\boldsymbol{T}_{n_{k+1}}(\boldsymbol{x}_0)\}$的极限，从而$\boldsymbol{T}(\boldsymbol{b}) \in B$. 依此类推，对任意正整数 j，我们有 $\boldsymbol{T}_j(\boldsymbol{b}) \in B$. 由于 B 是有界闭集，所以$\{\boldsymbol{T}_j(\boldsymbol{b}), j \geqslant 1\}$必有子序列且有极限，并且属于 B. 故由定义 2.2.13 知 $\boldsymbol{b} \in A_{\boldsymbol{T}}(B)$. 由于 $\boldsymbol{b}$ 是 B 中任一元素，故(2.2.29)式成立. ▌

在此顺便指出，若变换 $\boldsymbol{T}$ 不是连续的，(2.2.29)式不一定总是成立的. 换言之，吸引元 B 的吸引域可以不包括 B.

性质 2.2.15　若两个不同的吸引元是极限点，或是极限环，则它们的吸引域无公共元.

性质 2.2.16　若 B_1 和 B_2 是两个吸引元，并且 $B_1 \subset B_2$，则 $A_{\boldsymbol{T}}(B_1) \subset A_{\boldsymbol{T}}(B_2)$.

以上两个性质是显然的，证明从略.

如前所述，线性系统的吸引元只有极限点和极限环. 线性系统中的变换 $\boldsymbol{T}$ 又总是连续的，故由性质 2.2.15，线性系统的不同吸引元的吸引域是不相交的. 对于非线性系统的非极限点和非极限环而言，性质 2.2.15 的结论不一定适用. 例如在例 2.2.9 中，$D=[0,1]$是系统(2.2.13)的最大的吸引元，其他吸引元都是它的子集. 由性质 2.2.15 和性质 2.2.16 知，吸引元$[0,1]$的吸引域是其自身. 其他吸引元，如极限点$\{0\}$，极限环$\left\{\frac{1}{3}, \frac{2}{3}\right\}$，其吸引域都是$[0,1]$的子集.

下面讨论几个例子，用以说明吸引域的概念.

我们首先讨论线性系统(2.1.1). 在性质 2.1.2 中(i)的情况，$\{a_0/(1-a_1)\}$是唯一的吸引元，它的吸引域是全实轴. 在性质 2.1.2 中(ii)的情况，极限点$\{a_0/(1-a_1)\}$的吸引域为其自身. 在性质 2.1.2 中(iv)

的情况，每个实数都是极限点，而每个极限点的吸引域为其自身. 在性质 2.1.2中(v)的情况，每个极限环$\{-x_0, x_0\}(x_0 \neq a_0/2)$的吸引域都是其自身，极限点$\{a_0/2\}$的吸引域也是其自身. 根据§2.1的讨论，不难分析清楚任何线性系统的吸引元和吸引域. 实际上，当仅仅讨论线性系统时，没有必要引入吸引元和吸引域等概念.

下面讨论例 2.2.8 至例 2.2.12 所述的动态系统吸引元的吸引域.

如例 2.2.8 所述，动态系统(2.2.11)有唯一的单纯吸引元$[\alpha,\beta]$. 不难看出，无论从怎样的初始值 x_0 出发，由(2.2.11)式产生的迭代序列$\{x_t\}=\{T_t(x_0)\}$为在$[\alpha,\beta]$中的稠密子集. 于是由定义 2.2.13 可知 $x_0 \in A_T([\alpha,\beta])$. 直观地说，系统(2.2.11)将全实轴吸引到区间$[\alpha,\beta]$上. 类似地，系统(2.2.12)有唯一的单纯吸引元$[\log\alpha,\log\beta]$，它将全实轴吸引到$[\log\alpha,\log\beta]$上.

在例 2.2.9 中，系统(2.2.13)有吸引元$[0,1]$和可数多个极限环. 吸引元$[0,1]$是最大的，吸引域为其自身. 有趣的是，系统(2.2.13)的极限环的吸引域虽然不太容易给出，但都是 Lebesgue 零测集. 这将在第 8 章讨论.

在例 2.2.10 中，$\{0\}$是系统(2.2.15)的极限点. 虽然它的吸引域包括了所有的正无理数和零，但要给出此吸引域的精确表述，目前还是不可能的，因为它涉及悬而未决的 $3x+1$ 问题. 同样，极限环$\{1,4,2\}$的吸引域也是无法精确表述的. 这又一次看到，讨论非线性系统的吸引元和吸引域是何等困难.

对例 2.2.11 中所述的动态系统的吸引域，事实上在例 2.2.11 中的讨论已包括了如下的结论：

(i) 当 $0<\alpha\leqslant 1$ 时，$\{0\}$是系统(2.2.17)的极限点，它的吸引域为$[0,\infty)$，即 $A_T(\{0\})=[0,\infty)$.

(ii) 当 $\alpha>1$ 时，$A_T(\{0\})=\{0\}$，即极限点$\{0\}$只吸引其自身. 而 $A_T(\{(\alpha-1)^{1/2}\})=(0,\infty)$，即极限点$\{(\alpha-1)^{1/2}\}$吸引$(0,\infty)$中的每个元素，但不吸引 0.

例 2.2.12 所述动态系统的吸引集由 $D=[0,1]\times[0,1]$中的对角线元素所组成，而且每个元素$(c,c)^\tau$ 都是极限点. 由(2.2.24)式有

$$\frac{2}{3}x_t^2+\frac{1}{3}x_{t-1}^2=\frac{2}{3}\cdot\frac{1}{2}(x_{t-1}^2+x_{t-2}^2)+\frac{1}{3}x_{t-1}^2$$
$$=\frac{2}{3}x_{t-1}^2+\frac{1}{3}x_{t-2}^2.$$

由上式递推可知：

$$\frac{2}{3}x_t^2+\frac{1}{3}x_{t-1}^2=\frac{2}{3}x_{t-1}^2+\frac{1}{3}x_{t-2}^2=\cdots=\frac{2}{3}x_0^2+\frac{1}{3}x_{-1}^2.$$

若取 $x_0=b$，$x_{-1}=a$，则

$$\frac{2}{3}x_t^2+\frac{1}{3}x_{t-1}^2=\frac{2}{3}b^2+\frac{1}{3}a^2\quad(t=0,1,\cdots).$$

由例 2.2.12 的讨论还可知：

$$\lim_{t\to\infty}x_t^2=\frac{2}{3}b^2+\frac{1}{3}a^2.$$

联合以上两式可知，被极限点 $(c,c)^T$ 所吸引的初始值为 $(x_0,x_{-1})^T=(b,a)^T$，当且仅当

$$\frac{2}{3}b^2+\frac{1}{3}a^2=c.$$

满足上式的 $(b,a)^T\in D$ 为一椭圆轨道的一部分，见图 2.2.4.

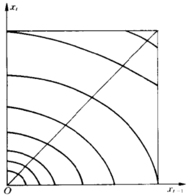


图 2.2.4　系统(2.2.24)的吸引域
图中的每条椭圆轨道为
相应的极限点 $(c,c)^T$ 的吸引域

由上述讨论可见，吸引域有不同的分类. 例如，有的吸引域 $A_T(B)=B$ 即只吸引自身，有的吸引域 $A_T(B)$ 是 Lebesgue 零测集，有的是 Lebesgue 正测集，等等. 而对高维系统，其分类更为复杂. 由于本书不讨论这些内容，所以关于吸引元和吸引域的介绍仅限于上述的初等概念性知识. 与本书关系密切的内容是：动态系统的吸引元吸引其吸引域的速度，特别是其吸引集的有界性以及吸引的速度问题. 这就是 § 2.3 中将要讨论的动态系统的稳定性问题.

§ 2.3　动态系统的稳定性

这一节将讨论动态系统的吸引元的吸引能力和速度. 在 § 2.2 中介绍了动态系统的吸引元，它们描述了动态系统轨道的所有可能的极限行为的分类特征. 此外还介绍了吸引元的吸引域，它们描述了每个吸引元的

吸引范围. 虽然这些概念能够刻划动态系统的重要结构特征，但还远远不够充分. 本节要讨论的吸引元的吸引能力就是要描述不同吸引域的范围的大小，可以形象地称为**吸引能力**. 比如，在§2.2中曾提到，有的吸引域只含一个元素，有的含有可列多个元素. 有的吸引域所含元素虽多，却是 Lebesgue 零测集，还有的吸引域为全空间. 这表明不同的吸引元的吸引能力有很大差别，当然应加以分类区别. 另一方面，吸引元在吸引其吸引域的元素时，还有吸引速度的问题. 换言之，从某一 $\boldsymbol{x}_0$ 出发，若动态系统(2.2.1)的轨道 $\boldsymbol{x}_t=\boldsymbol{T}_t(\boldsymbol{x}_0)$ 向某一吸引元收敛，其收敛速度也是描述动态系统特征的重要内容. 以上两方面的内容，都属于动态系统的稳定性问题.

稳定性概念是刻划动态系统特征的重要方面. 对不同类型的稳定性的定义和特征的研究，有相当丰富的内容(见文献[106]). 本书不准备对这方面的内容详细介绍，而是着重讨论与本书有关的内容.

定义 2.3.1 如果由动态系统(2.2.2)产生的迭代序列 $\{\boldsymbol{T}_t(\boldsymbol{x}_0), t\geqslant 1\}$ 满足

$$\limsup_{t\to\infty} \|\boldsymbol{T}_t(\boldsymbol{x}_0)\| < \infty, \tag{2.3.1}$$

则称系统(2.2.2)是 **Lagrange 稳定的**.

实质上，Lagrange 稳定的动态系统是指无发散轨道的动态系统.

定义 2.3.2 如果由动态系统(2.2.2)产生的迭代序列 $\{\boldsymbol{T}_t(\boldsymbol{x}_0), t\geqslant 1\}$ 满足

$$\lim_{t\to\infty} \|\boldsymbol{T}_t(\boldsymbol{x}_0)\| = \infty, \tag{2.3.2}$$

则称系统(2.2.2)**在 $\boldsymbol{x}_0$ 处发散**. 若(2.3.2)式对每个 $\boldsymbol{x}_0\in D$ 都成立，则称系统(2.2.2)是**整体发散的**.

例如，在线性系统(2.1.1)中，对性质 2.1.2 中的(ii)，$x_0\neq a_0/(1-a_1)$ 是该系统的发散点. 对性质 2.1.2 中的(iii)，系统(2.1.1)是整体发散的. 在本书中，我们对有发散现象的动态系统不进行更多地讨论.

由定义 2.3.1 给出的稳定性太广义了. 例如对下面的动态系统

$$\begin{cases} x_t = x_{t-1} \quad (t\geqslant 1); \\ x_0 \in R^1, \end{cases} \tag{2.3.3}$$

按定义 2.3.1，它是 Lagrange 稳定的. 显然，R^1 中每个元素都是系统(2.3.3)的不动点，且每个不动点除自身外，不吸引任何其他点. 在时间序列分析中，与(2.3.3)式相应的一元 LAR 模型(详见本书例 4.2.2)

$$x_t = x_{t-1} + \varepsilon_t$$

是没有平稳解的. 因此, 以上非常广义的稳定性概念与本书所研究的时间序列模型的关系并不密切, 我们对其不作进一步讨论了. 下面讨论与本书关系密切的稳定性概念.

定义 2.3.3　设 $L \subset R^m$ 是一有界集, $\{\boldsymbol{T}_t(\boldsymbol{x}_0), t \geqslant 1\}$ 是动态系统(2.2.2)以 $\boldsymbol{x}_0$ 为初始值所产生的迭代序列, $r_t(\boldsymbol{x}_0, L)$ 表示 $\boldsymbol{T}_t(\boldsymbol{x}_0)$ 到集合 L 的距离, 并由(2.2.28)式所定义.

(i) 如果

$$\lim_{t\to\infty} r_t(\boldsymbol{x}_0, L) = 0, \quad \forall\, \boldsymbol{x}_0 \in D, \tag{2.3.4}$$

则称动态系统(2.2.2)在 L 处为*整体稳定的*.

(ii) 如果(2.3.4)式的收敛具有指数速度, 即存在正数 $\rho < 1$ 和 K_1、K_2, 使得

$$r_t(\boldsymbol{x}_0, L) \leqslant \rho^t [K_1 r(\boldsymbol{x}_0, L) + K_2], \quad \forall\, \boldsymbol{x}_0 \in D, \tag{2.3.5}$$

其中 $r(\boldsymbol{x}_0, L)$ 表示 $\boldsymbol{x}_0$ 到 L 的距离, 则称动态系统(2.2.2)在 L 处为*几何速度稳定的*.

定义 2.3.3 中的稳定性均指在有界集 L 处稳定. 在后面叙述中若不明确强调有界集 L 的特殊性时, 我们省略“在 L 处”三字. 以上定义的稳定性均指从任何 $\boldsymbol{x}_0$ 出发, 动态系统(2.2.2)的轨道 $\boldsymbol{T}_t(\boldsymbol{x}_0)$ 必趋向于有界集 L. 按 § 2.2 中吸引集的定义, 可以说 L 是动态系统(2.2.2)的一个广义吸引集, 而且其吸引域为全体可能的 $\boldsymbol{x}_0 \in D$, 故称*整体稳定*. 在以后叙述中, 为了简略, 除在强调局部稳定时, “稳定”一词总指**整体稳定**.

在定义 2.3.3 中, 如果动态系统(2.2.2)在 L 处为几何速度稳定, 则可以说, L 是具有指数速度的吸引集. 在文献[30]中曾用下式:

$$r_t(\boldsymbol{x}_0, L) \leqslant K\rho^t, \quad \forall\, \boldsymbol{x}_0 \in D \tag{2.3.6}$$

代替(2.3.5)式, 以定义几何速度稳定. 它相当于在(2.3.5)式中取 $K_1 = 0$, $K_2 = K$. 这样定义的几何速度稳定太严了. 比如, 对最简单的一元一阶压缩变换 $T(x) = \rho x$, 其中 $0 < |\rho| < 1$, 相应的迭代序列为 $\{\boldsymbol{T}_t(\boldsymbol{x}_0) = \rho^t \boldsymbol{x}_0, t \geqslant 1\}$, 相应的吸引集为 $L(T) = \{0\}$, 但是

$$r_t(\boldsymbol{x}_0, \{0\}) = |\rho|^t |\boldsymbol{x}_0|, \quad \forall\, \boldsymbol{x}_0 \in D.$$

由此可见, $T(x) = \rho x$ 相应的动态系统(2.2.2)的吸引集不是(2.3.6)式意义下的几何速度稳定的.

实际上, 当 $D = R^m$ 时, (2.3.6)式意味着: 对每个 $t \geqslant 1$, $r(\boldsymbol{T}_t(\boldsymbol{x}_0), L)$

$\leqslant K$,从而 $\boldsymbol{T}_1(\boldsymbol{x}_0)=\boldsymbol{T}(\boldsymbol{x}_0)$ 必有界,即 $\boldsymbol{T}(\boldsymbol{x})$ 是有界函数. 这也说明由(2.3.6)式定义的几何速度稳定只适用于有界变换 $\boldsymbol{T}$ 相应的动态系统. 所以我们不以(2.3.6)式作为几何速度稳定的标准,而以(2.3.5)式来定义几何速度稳定.

另一方面,考虑变换

$$T(x)=\rho x+a, \tag{2.3.7}$$

其中 $0<\rho<1$, $a\neq 0$,易见

$$\begin{aligned} x_t &= T(x_{t-1})=\rho x_{t-1}+a=\cdots \\ &= a+a\rho+\cdots+a\rho^{t-1}+\rho^t x_0 \\ &= a/(1-\rho)+\rho^t[x_0-a/(1-\rho)]. \end{aligned} \tag{2.3.8}$$

对变换(2.3.7),吸引集 $L(T)=\{a/(1-\rho)\}$. 于是由(2.3.8)式,有

$$\begin{aligned} r_t(x_0,\{a/(1-\rho)\}) &= \rho^t|x_0-a/(1-\rho)| \\ &= \rho^t r(x_0,\{a/(1-\rho)\}), \quad \forall\, x_0\in R^1. \end{aligned}$$

上式说明,变换(2.3.7)满足(2.3.5)式,且相应于 $K_1=1$ 和 $K_2=0$ 的情况. 在(2.3.5)式中不限定 $K_2=0$,是为了使定义内涵更广些.

还需指出,当(2.3.5)式成立时,必然存在正整数 $\rho_1<1$,且 $\rho_1>\rho$,使得

$$\lim_{t\to\infty}\rho_1^{-t}r_t(\boldsymbol{x}_0,L)=0, \quad \forall\, x_0\in D. \tag{2.3.9}$$

反之,当(2.3.9)式成立时,不一定保证(2.3.5)式成立. 例如考虑如下变换:

$$T(x)=\begin{cases} x-1, & 当\ x\geqslant 0; \\ -1, & 当\ x<0, \end{cases}$$

其相应的动态系统为

$$x_t=\begin{cases} x_{t-1}-1, & 当\ x_{t-1}\geqslant 0; \\ -1, & 当\ x_{t-1}<0. \end{cases} \tag{2.3.10}$$

显然,动态系统(2.3.10)的吸引集 $L(T)=\{-1\}$,而且从任何 $\boldsymbol{x}_0$ 出发,进行有限步迭代后,必到达-1. 故(2.3.9)式成立. 但是,当取 $\boldsymbol{x}_0$ 为正整数 n 时,易见

$$r_t(n,\{-1\})=n+1-t \quad (t=1,2,\cdots,n),$$
$$r(n,\{-1\})=n+1.$$

若(2.3.5)式成立,则有

$$(n+1-t)\leqslant\rho^t(n+1+c).$$

于是有

$$\rho^t \geqslant \frac{n+1-t}{n+1+c}, \quad \forall\, t=1,2,\cdots,n;\ n=1,2,\cdots. \qquad (2.3.11)$$

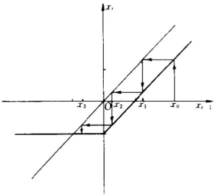


图 2.3.1　动态系统(2.3.10)的迭代示意图

易见不可能存在正数 $\rho<1$,使得(2.3.11)式成立. 于是导致矛盾!若以(2.3.9)式代替(2.3.5)式,则可定义另一种更广义的几何速度稳定. 但是本书不讨论此类定义.

显然,定义 2.3.3 下的稳定性比定义 2.3.1 更强. 根据 §2.2 吸引元的定义可知,定义 2.3.3 下的稳定性保证了动态系统(2.2.2)有一定的吸引性质.

性质 2.3.4　动态系统(2.2.2)在有界集 $L\subset R^m$ 上为稳定的,当且仅当动态系统(2.2.2)的吸引集是 L 的子集.

此性质的结论是显然的,证明从略.

定义 2.3.5　设 L_1 是 R^m 中的有界集,D_1 是动态系统(2.2.2)中 D 的子集,且 D_1 至少包含 R^m 中的一个 m 维立方体. 如果

$$\lim_{t\to\infty} r_t(\boldsymbol{x}_0, L_1) = 0, \quad \forall\, \boldsymbol{x}_0 \in D_1, \qquad (2.3.12)$$

则称动态系统(2.2.2)在 L_1 处为*局部稳定的*.

对照定义 2.3.5 和定义 2.3.3,我们作如下两点说明:其一是,满足(2.3.12)式的局部稳定性是相对于 $D_1\subset D$ 且 $D_1\neq D$ 而言的. 如果 $D_1=D$,则局部稳定就是整体稳定. 若已知动态系统(2.2.2)为局部稳定的,则将(2.2.2)式改写为

$$\begin{cases} \boldsymbol{x}_t = \boldsymbol{T}(\boldsymbol{x}_{t-1}); \\ \boldsymbol{x}_0 \in D_1. \end{cases} \qquad (2.3.13)$$

可见,动态系统(2.3.13)与(2.2.2)的区别仅在于初始值 $\boldsymbol{x}_0$ 的取值范围不同. 于是,相对于动态系统(2.3.13)而言,在 L_1 处的局部稳定性即是整体稳定性. 另外,动态系统(2.3.13)还可以是几何速度稳定的. 但是在考查动态系统(2.2.2)的稳定性时,如果动态系统(2.3.13)是几何速度稳定的,则应冠以"局部"一词,即称动态系统(2.2.2)在 L_1 处为局部几何速度稳定的.

其二是，在定义 2.3.5 中，区域 D_1 被有界集 L_1 所吸引，这很像 § 2.2 中的吸引域与吸引元的关系. 但是在定义 2.3.5 中，特别强调了 D_1 至少包含 R^m 中一个 m 维立方体，这就保证了 D_1 是不降维子集，而且 Lebesgue 测度是正的. 按照这样的定义，如果 § 2.2 中的吸引元的吸引域为有限点集. 或者为 Lebesgue 零测集，则它们就不可能成为定义 2.3.5 中的 D_1. 或者说，在这样的吸引元处，动态系统(2.2.2)不具有局部稳定性.

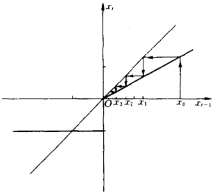


图 2.3.2　动态系统(2.3.14)的迭代示意图

关于局部稳定性还可以采用其他方式定义. 比如，设 $L_1 \subset R^m$ 是有界集，它能吸引区域 D_1，即满足(2.3.12)式. 如果 D_1 还包含了 L_1 的某个邻域，则称动态系统(2.2.2)在 L_1 处具有局部稳定性. 这样定义的稳定性，对于 $\boldsymbol{T}$ 为连续可微的变换比较适用. 当 $\boldsymbol{T}$ 不连续时，这样的定义过于狭隘了. 例如，对如下的动态系统：

$$x_t = \begin{cases} \dfrac{1}{2}x_{t-1}, & \text{当 } x_{t-1} > 0; \\ -1, & \text{当 } x_{t-1} \leqslant 0, \end{cases} \tag{2.3.14}$$

易见{0}是动态系统(2.3.14)的一个极限点. $L_1 = \{0\}$ 为一单点吸引元. 又易见，它的吸引域 $A_T(L_1) = (0, \infty)$. 这里有两点很特别，其一，L_1 的吸引域不包含吸引元自身. 换言之，L_1 吸引$(0,\infty)$中的每个元素，却不吸引它自身，因为(2.3.14)式以 $x_0 = 0$ 为初始值的轨道 $T_t(0) = -1$；其二，按定义 2.3.5，动态系统(2.3.14)在 $L_1 = \{0\}$ 处是局部稳定的. 实际上还是几何速度局部稳定的. 但是，若采用上述定义，{0}至少吸引{0}的一个邻域才可称为局部稳定. 可见，在上述定义下，吸引元 $L_1 = \{0\}$ 不是局部稳定的.

由于本书力图包括不连续的系统在内而加以讨论，所以我们按定义 2.3.5 给出如下稳定吸引元的定义：

定义 2.3.6　动态系统(2.2.2)的吸引元 B 被称为稳定的，如果动态系统(2.2.2)在 B 处为局部稳定的. 否则，称其为非稳定的吸引元.

按照定义 2.3.5 下面的补充说明，对每个吸引元，除了区别出稳定和

非稳定之外，还可按其吸引的速度分为几何速度与非几何速度. 这些细微的分类是极易分别定义的，这里就不一一罗列了. 与本书所讨论的非线性时间序列关系最为密切的是定义 2.3.3 中的具有几何速度的整体稳定系统. 这样的系统相应于定义 2.3.3 中的有界集 L. 按照性质 2.3.4，它包括了动态系统(2.2.2)的所有吸引元. 虽然按照定义 2.3.5 和定义 2.3.6 可以分别研究动态系统(2.2.2)的每个吸引元的吸引性质，但是，本书以后的内容并不涉及这样更细致的动态系统的局部特征，因此也就不再深入讨论有关动态系统的局部稳定问题了.

在文献[28]中，对于具有压缩性质的动态系统研究得较早，也受到较多的注意，而且这种动态系统与熟知的线性系统关系也密切. 为此，我们特别给出以下的定义和某些性质：

定义 2.3.7　动态系统(2.2.2)称为依范数 $\|\cdot\|_v$ 压缩的，如果存在正数 $0\leqslant\rho<1$，使得对任意的 $\boldsymbol{x}\in D$，有

$$\|\boldsymbol{T}(\boldsymbol{x})\|_v \leqslant \rho\|\boldsymbol{x}\|_v. \tag{2.3.15}$$

定理 2.3.8　如果动态系统(2.2.2)是依范数 $\|\cdot\|_v$ 压缩的，则系统(2.2.2)在零点几何速度稳定，从而系统(2.2.2)的吸引集为只包含零点的单点集.

证明　由于变换 $\boldsymbol{T}$ 是依范数 $\|\cdot\|_v$ 压缩的，故有

$$\begin{aligned}\|\boldsymbol{T}_n(\boldsymbol{x}_0)\|_v &= \|\boldsymbol{T}(\boldsymbol{T}_{n-1}(\boldsymbol{x}_0))\|_v \\ &\leqslant \rho\|\boldsymbol{T}_{n-1}(\boldsymbol{x}_0)\|_v \\ &\leqslant \cdots \\ &\leqslant \rho^n\|\boldsymbol{T}_0(\boldsymbol{x}_0)\|_v \\ &= \rho^n\|\boldsymbol{x}_0\|_v, \quad \forall x_0\in D.\end{aligned} \tag{2.3.16}$$

注意到范数 $\|\cdot\|_v$ 与范数 $\|\cdot\|$ 是等价的，即存在 c_1、$c_2>0$，使得

$$c_1\|\boldsymbol{x}\| \leqslant \|\boldsymbol{x}\|_v \leqslant c_2\|\boldsymbol{x}\|, \ \forall \boldsymbol{x}\in D. \tag{2.3.17}$$

联合(2.3.16)和(2.3.17)两式，存在 c_1、$c_2>0$，使得

$$c_1\|\boldsymbol{T}_n(\boldsymbol{x}_0)\| \leqslant \|\boldsymbol{T}_n(\boldsymbol{x}_0)\|_v \leqslant c_2\rho^n\|\boldsymbol{x}_0\|.$$

记 $L=\{0\}$，则上式意味着

$$r_n(\boldsymbol{x}_0,L) \leqslant c\rho^n r(\boldsymbol{x}_0,L),$$

其中 $c=c_2/c_1$. 故依定义 2.3.3 知，系统(2.2.2)在零点为几何速度稳定的. ∎

下面考查一个简单的二维线性系统：

$$\begin{cases} \boldsymbol{x}_t = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} \boldsymbol{x}_{t-1}; \\ \boldsymbol{x}_0 \in R^2, \end{cases} \tag{2.3.18}$$

其中 $0<\lambda<1$. 记 $\boldsymbol{T}(\boldsymbol{x})=A\boldsymbol{x}$,其中

$$A = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}.$$

取 $\boldsymbol{x}=(0,x)^\tau$,则由(2.3.18)式有

$$\begin{aligned} \| \boldsymbol{T}(\boldsymbol{x}) \|^2 &= \| A\boldsymbol{x} \|^2 = \boldsymbol{x}^\tau A^\tau A\boldsymbol{x} \\ &= (0,x)\begin{pmatrix} \lambda & 0 \\ 1 & \lambda \end{pmatrix}\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}\begin{pmatrix} 0 \\ x \end{pmatrix} \\ &= (1+\lambda^2)x^2 = (1+\lambda^2)\| \boldsymbol{x} \|^2. \end{aligned}$$

由上式可得

$$\| \boldsymbol{T}(\boldsymbol{x}) \| = (1+\lambda^2)^{1/2} \| \boldsymbol{x} \| > \| \boldsymbol{x} \|.$$

故线性系统(2.3.18)不是依平方范数压缩的.

以上例子说明,对于线性系统(2.3.18),参数矩阵 A 的特征值的模都小于1,并不能保证它在平方范数下为压缩的. 试问,这一条件是否能保证线性系统

$$\begin{cases} \boldsymbol{x}_t = A\boldsymbol{x}_{t-1}; \\ \boldsymbol{x}_0 \in R^m, \end{cases} \tag{2.3.19}$$

依某范数为压缩呢? 下面的定理回答了这一问题.

定理 2.3.9 如果线性系统(2.3.19)中参数矩阵 A 的特征值的模都小于1,则必存在 R^m 中的一个范数 $\|\cdot\|_v$,使得线性系统(2.3.19)依范数 $\|\cdot\|_v$ 为压缩的.

证明 设 Q 为一正定对称方阵,对 $\boldsymbol{x}\in R^m$ 定义如下向量范数:

$$\| \boldsymbol{x} \|_Q = (\boldsymbol{x}^\tau Q\boldsymbol{x})^{1/2}. \tag{2.3.20}$$

由(2.3.19)式有

$$\| A\boldsymbol{x} \|_Q^2 = \boldsymbol{x}^\tau A^\tau QA\boldsymbol{x}. \tag{2.3.21}$$

由于矩阵 A 的特征值的模都小于1,根据文献[36]中的第19页或用归纳法,可以证明存在正定对称方阵 Q 和常数 $0<\lambda<1$,使得

$$A^\tau QA \leqslant \lambda^2 Q, \tag{2.3.22}$$

联合(2.3.21)和(2.3.22)两式得

$$\| A\boldsymbol{x} \|_Q \leqslant \lambda \| \boldsymbol{x} \|_Q,$$

从而完成定理的证明. ▌

推论 2.3.10 线性系统(2.3.19)在零点几何速度稳定,当且仅当参数矩阵 A 的特征值的模都小于 1. 当参数矩阵 A 具有如下形式:

$$A = \begin{pmatrix} a_1 & \cdots & a_{p-1} & a_p \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix}$$

时,A 的特征值的模都小于 1,当且仅当参数 $a_1,\cdots,a_p$ 满足

$$1 - a_1 z - \cdots - a_p z^p \neq 0, \quad |z| \leqslant 1. \tag{2.3.23}$$

证明是显然的,从略.

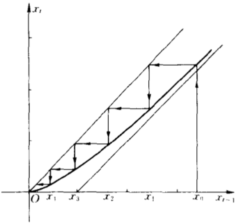


图 2.3.3 动态系统(2.3.24)的迭代示意图

注 2.3.11 对于线性系统(2.1.3),根据推论 2.3.10,不难验证,如果线性系统(2.1.3)在 $(\boldsymbol{I}-A)^{-1}\boldsymbol{a}$ 处稳定,则必为几何速度稳定. 但对于非线性动态系统(2.2.2),在 L 处稳定,未必一定有几何速度稳定. 其原因在于:对线性系统(2.1.3),有

$$\frac{\partial(A\boldsymbol{x}+\boldsymbol{a})}{\partial \boldsymbol{x}} = A,$$

而对非线性系统(2.2.2),有

$$\frac{\partial \boldsymbol{T}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{\Psi}(\boldsymbol{x})$$

为 $\boldsymbol{x}$ 的矩阵函数,而非常数矩阵. 例如对如下的一阶一维非线性系统

$$\begin{cases} x_t = \dfrac{x_{t-1}}{1+x_{t-1}} \quad (t \geqslant 1); \\ x_0 \in [0,\infty), \end{cases} \tag{2.3.24}$$

易见

$$\frac{\mathrm{d}T(\boldsymbol{x})}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x}{1+x}\right) = \frac{1}{(1+x)^2}.$$

不难验证,$x=0$ 是动态系统(2.3.24)的不动点,且动态系统(2.3.24)在零点稳定,但没有几何速度稳定性.

下面讨论如何判断非线性系统具有几何速度稳定性.

定理 2.3.12 如果动态系统(2.2.2)的变换 $\boldsymbol{T}(\boldsymbol{x})$ 在某向量范数 $\|\cdot\|_c$ 之下,满足

$$\| \boldsymbol{T}(\boldsymbol{x}) \|_v \leqslant \rho \| \boldsymbol{x} \|_v + c, \tag{2.3.25}$$

其中 $c \geqslant 0$ 和 $0 \leqslant \rho < 1$ 为常数，则动态系统(2.2.2)在有界集 $L = \{\boldsymbol{x} : \| \boldsymbol{x} \|_v \leqslant c/(1-\rho)\}$ 处为几何速度稳定的.

证明 由(2.3.25)式有

$$\begin{aligned}
\| \boldsymbol{x}_t \|_v &= \| \boldsymbol{T}_t(\boldsymbol{x}_0) \|_v = \| \boldsymbol{T}(\boldsymbol{T}_{t-1}(\boldsymbol{x}_0)) \|_v \\
&\leqslant \rho \| \boldsymbol{T}_{t-1}(\boldsymbol{x}_0) \|_v + c \leqslant \cdots \\
&\leqslant c + c\rho + c\rho^2 + \cdots + c\rho^{t-1} + \rho^t \| \boldsymbol{x}_0 \|_v \\
&= c/(1-\rho) + \rho^t \| \boldsymbol{x}_0 \|_v - \sum_{j=t}^{\infty} c\rho^j \\
&\leqslant c/(1-\rho) + \rho^t \| \boldsymbol{x}_0 \|_v .
\end{aligned}$$

于是，再由(2.2.28)式和范数的等价性(即(2.3.17)式)以及上式可得

$$\begin{aligned}
r_t(\boldsymbol{x}_0, L) &= \inf_{y \in L} \| \boldsymbol{x}_t - \boldsymbol{y} \| \\
&= \inf_{\| \boldsymbol{y} \|_v \leqslant \frac{c}{1-\rho}} \| \boldsymbol{x}_t - \boldsymbol{y} \| \\
&\leqslant \left\| \boldsymbol{x}_t - \frac{c/(1-\rho)}{c/(1-\rho) + \rho^t \| \boldsymbol{x}_0 \|_v} \boldsymbol{x}_t \right\| \\
&\leqslant \frac{c_1^{-1} \rho^t \| \boldsymbol{x}_0 \|_v}{c/(1-\rho) + \rho^t \| \boldsymbol{x}_0 \|_v} \| \boldsymbol{x}_t \|_v \\
&\leqslant c_1^{-1} \rho^t \| \boldsymbol{x}_0 \|_v \leqslant \frac{c_2}{c_1} \rho^t \| \boldsymbol{x}_0 \| \\
&= \frac{c_2}{c_1} \rho^t \left(\inf_{\| \boldsymbol{y} \|_v \leqslant \frac{c}{1-\rho}} \| \boldsymbol{x}_0 - \boldsymbol{y} + \boldsymbol{y} \| \right) \\
&\leqslant \rho^t \left[\frac{c_2}{c_1} r(\boldsymbol{x}_0, L) + c_2 c/(1-\rho) c_1 \right].
\end{aligned}$$

由于上式中的 $\boldsymbol{x}_0 \in D$ 是任意的，这表明(2.3.5)式成立，从而动态系统(2.2.2)在 L 处为几何速度稳定的. ▌

定理 2.3.13 如果动态系统(2.2.2)中的变换 $\boldsymbol{T}(\boldsymbol{x})$ 满足

$$\boldsymbol{T}(\boldsymbol{x}) = A\boldsymbol{x} + \boldsymbol{H}(\boldsymbol{x}), \tag{2.3.26}$$

并假定：

(i) $\rho(A) < 1$, (2.3.27)

其中 $\rho(A) = \max\{|\lambda| : \lambda$ 是 A 的特征值$\}$是矩阵 A 的谱半径.

(ii) 对任意的 $K > 0$，

$$\sup_{\| \boldsymbol{x} \| \leqslant K} \| \boldsymbol{H}(\boldsymbol{x}) \| < \infty, \tag{2.3.28}$$

且当 $\|x\|\to\infty$ 时，

$$\frac{\|H(x)\|}{\|x\|}\to 0. \tag{2.3.29}$$

则动态系统(2.2.2)必在 R^m 中的某有界集上为几何速度稳定的.

证明 由(2.3.27)式和定理 2.3.9 的证明知，存在一个 m 阶正定对称方阵 Q，使得在(2.3.20)式定义的向量范数 $\|\cdot\|_Q$ 下，有

$$\|Ax\|_Q\leqslant\rho\|x\|_Q,$$

其中 $0<\rho<1$ 为常数. 这样，由(2.3.26)式有

$$\begin{aligned}&\|T(x)\|_Q\\&\quad=\|Ax+H(x)\|_Q\\&\quad\leqslant\|Ax\|_Q+\|H(x)\|_Q\\&\quad\leqslant\rho\|x\|_Q+(\|H(x)\|_Q/\|x\|_Q)\|x\|_Q.\end{aligned}$$

又由向量范数的等价性，存在常数 $0<c_1\leqslant c_2<\infty$，使得

$$\|T(x)\|_Q\leqslant\rho\|x\|_Q+(c_2/c_1)(\|H(x)\|/\|x\|)\|x\|_Q. \tag{2.3.30}$$

由(2.3.29)式知，存在常数 $\rho<\gamma<1$ 和 $K>0$，使得当 $\|x\|\geqslant K$ 时，

$$\frac{c_2}{c_1}\cdot\frac{\|H(x)\|}{\|x\|}\leqslant\gamma-\rho. \tag{2.3.31}$$

再由(2.3.28)式，存在常数 $c>0$，使得

$$\sup_{\|x\|\leqslant K}\|H(x)\|\leqslant c<\infty,$$

于是联合上式和(2.3.30)以及(2.3.31)，便有

$$\|T(x)\|_Q\leqslant\gamma\|x\|_Q+c.$$

根据定理 2.3.12 知，由变换(2.3.26)确定的动态系统在有界集 $L=\{x:\|x\|_Q\leqslant c/(1-\rho)\}$ 处为几何速度稳定的. ▌

推论 2.3.14 设动态系统(2.2.2)在 L 处为几何速度稳定的，L_1 为任一有界集且 $L_1\subset L$，则动态系统(2.2.2)在 L_1 上也是几何速度稳定的.

推论的结论是显然的，证明从略.

注 2.3.15 在定理 2.3.13 相同的条件下，若将(2.3.29)式代之以存在正数 $\lambda'<1$，且 $\rho+\lambda'<1$，使得

$$\limsup_{\|x\|\to\infty}\left(\frac{c_2}{c_1}\right)\frac{\|H(x)\|}{\|x\|}=\lambda', \tag{2.3.32}$$

则定理 2.3.13 的结论仍成立.

需要指出的是：定理 2.3.13 虽然不是动态系统(2.2.2)为几何速度

稳定的充分必要条件，但从证明过程可以看出，它们是接近必要的．特别是条件(2.3.32)更接近必要性．

定理 2.3.16　设动态系统(2.2.2)中的变换 $\boldsymbol{T}(\boldsymbol{x})$，经 n 步迭代后，满足：

(i) $\|\boldsymbol{T}_j(\boldsymbol{x}_0)\|_v \leqslant c_1\|\boldsymbol{x}_0\|_v + c_2 \quad (j=1,2,\cdots,n-1)$，　　(2.3.33)

其中 $c_1\geqslant 0, c_2\geqslant 0$ 为常数，$\|\cdot\|_v$ 为某适当的向量范数．

(ii) $\boldsymbol{T}_n(\boldsymbol{x})$ 满足

$$\boldsymbol{T}_n(\boldsymbol{x}) = A\boldsymbol{x} + \boldsymbol{H}(\boldsymbol{x}), \tag{2.3.34}$$

且参数矩阵 A 和变换 $\boldsymbol{H}(\boldsymbol{x})$ 满足定理 2.3.13 的条件(i)和(ii)，或者满足(2.3.25)式，其中向量范数 $\|\cdot\|_v$ 与(2.3.33)式中的相同．

则动态系统必在某有界集 $L\subset R^m$ 处为几何速度稳定的．

证明　令 $\boldsymbol{y}_t=\boldsymbol{x}_{tn}$，则由(2.3.34)式知，$\{\boldsymbol{y}_t\}$ 满足如下动态系统：

$$\begin{cases}\boldsymbol{y}_t = A\boldsymbol{y}_{t-1} + \boldsymbol{H}(\boldsymbol{y}_{t-1}) \quad (t\geqslant 1);\\ \boldsymbol{y}_0 \in D.\end{cases} \tag{2.3.35}$$

由定理 2.3.13，存在 R^m 中的有界集 L 和常数 $0\leqslant\gamma<1, K_1、K_2\geqslant 0$(以下各式中出现，含义可不同)，使得

$$r_t(\boldsymbol{y}_0,L) \leqslant \gamma^t[K_1 r(\boldsymbol{y}_0,L) + K_2],$$

从而

$$r_{tn}(\boldsymbol{x}_0,L) \leqslant \gamma^t[K_1 r(\boldsymbol{x}_0,L) + K_2].$$

再令 $\boldsymbol{z}_t=\boldsymbol{x}_{t,n+1}$，同理可证

$$r_t(\boldsymbol{z}_0,L) \leqslant \gamma^t[K_1 r(\boldsymbol{z}_0,L) + K_2],$$

从而

$$r_{t,n+1}(\boldsymbol{x}_0,L) \leqslant \gamma^t[K_1 r(\boldsymbol{x}_1,L) + K_2].$$

依此类推，可证明对任何 $j<n$，有

$$r_{t,n+j}(\boldsymbol{x}_0,L) \leqslant \gamma^t[K_1 r(\boldsymbol{x}_j,L) + K_2]. \tag{2.3.36}$$

令 $\rho=\gamma^{1/n}$，由(2.3.33)式和向量范数的等价性，不难验证，对任意的 $j<n$，取 $c_1'=c_1/\rho^n$，$c_2'=c_2/\rho^n$，则有

$$r(\boldsymbol{x}_j,L) \leqslant \rho^j[c_1' r(\boldsymbol{x}_0,L) + c_2'].$$

再注意到，对于任意的 $t>n$，必存在 $k>0$ 和 $j\geqslant 0$，使得 $t=kn+j$，从而由(2.3.36)式可得

$$\begin{aligned} r_t(\boldsymbol{x}_0,L) &= r_{k,n+j}(\boldsymbol{x}_0,L) \\ &\leqslant \gamma^k[K_1 r(\boldsymbol{x}_j,L) + K_2] \end{aligned}$$

$$
\begin{aligned}
&\leqslant (\gamma^{\frac{1}{n}})^{nk}\{K_1\rho^j[c_1' r(\boldsymbol{x}_0,L) + c_2'] + K_2\} \\
&\leqslant \rho^{nk+j}[K_1c_1' r(\boldsymbol{x}_0,L) + K_1c_2' + \rho^{-j}\cdot K_2] \\
&= \rho^t[K_1' r(\boldsymbol{x}_0,L) + K_2'],
\end{aligned}
$$

其中 $K_1'=K_1c_1'$，$K_2'=K_1c_2'+\rho^{-j}\cdot K_2$. 上式即为(2.3.5)式. 于是定理得证. ▌

以上我们讨论了动态系统(2.2.2)在某有界集 $L\subset R^m$ 处几何速度稳定的条件. 但是，对某些动态系统，往往很难判断它们是否满足上述定理或推论的条件. 不过，对一维连续动态系统，判断其在极限点和极限环处的几何速度稳定有如下准则.

定理 2.3.17　考虑如下一阶一维连续动态系统：

$$
\begin{cases} x_t = \varphi(x_{t-1}), \\ x_0 \in R^1. \end{cases} \tag{2.3.37}
$$

若 $\varphi(0)=0$，$\varphi(x)$有连续的导数 $\varphi'(x)$，且 $|\varphi'(x)|\leqslant\rho<1$，则动态系统(2.3.37)在零点是几何速度稳定的.

证明　由定理的条件可得：

$$
|\varphi(x)| = |\varphi(x) - \varphi(0)| \leqslant \rho|x|, \quad \forall\, x\in R^1,
$$

即动态系统(2.3.37)是压缩的. 从而由定理 2.3.8 即可得证. ▌

推论 2.3.18　若动态系统(2.3.37)中的变换 φ 有连续的导数，且当 $x\in[a-\delta_1, a+\delta_2]$($\delta_1>0, \delta_2>0$)时，

$$
\varphi(a) = a, \quad |\varphi'(x)| < 1,
$$

则动态系统(2.3.37)在 a 点处为几何速度稳定的.

证明是显然的，从略.

注 2.3.19　由定义 2.3.5 知，推论 2.3.18 中的点集$\{a\}$实际上是动态系统(2.3.37)的局部稳定极限点.

推论 2.3.20　若动态系统(2.3.37)中的变换 $\varphi(x)$有周期解 $a_1,\cdots,a_q$，即

$$
\varphi(a_i) = a_{i+1} \quad (i=1,\cdots,q), \quad a_{q+1} = a_1,
$$

且 $\varphi(x)$有连续的导数 $\varphi'(x)$，并满足

$$
|\varphi'(a_1)\cdots\varphi'(a_q)| < 1,
$$

则$\{a_1,\cdots,a_q\}$是动态系统(2.3.37)的稳定极限环.

证明也是显然的，从略.

例 2.3.21　考虑一阶一维动态系统

$$\begin{cases} x_t = \alpha x_{t-1} + |x_{t-1}|^{\beta}\mathrm{sign}(x_{t-1}); \\ x_0 \in R^1. \end{cases} \tag{2.3.38}$$

式中 α 和 β 满足 $|\alpha|<1, 0<\beta<1$. 由(2.3.38)式可知,变换 $\varphi(x)=\alpha x+|x|^{\beta}\mathrm{sign}(x)$ 满足定理 2.3.13 的条件,从而动态系统(2.3.38)在某有界集处为几何速度稳定的. 此动态系统的吸引集为三个不动点,即 $\{0\}$、$\{-(1-\alpha)^{-1/(1-\beta)}\}$ 和 $\{(1-\alpha)^{-1/(1-\beta)}\}$. 由直接验证可知,$\{0\}$ 不是稳定的吸引元,利用推论 2.3.18 可得知其他两个不动点是局部几何稳定的.

例 2.3.22 考虑如下分段线性动态系统:

$$x_t = \begin{cases} \alpha_0 + \alpha_1 x_{t-1}, & \text{当 } x_{t-1} \geqslant 0; \\ \beta_0 + \beta_1 x_{t-1}, & \text{当 } x_{t-1} < 0, \end{cases} \tag{2.3.39}$$

$$x_0 \in R^1,$$

其中系数 α_1 和 β_1 满足如下条件:

$$\alpha_1 < 1,\ \beta_1 < 1,\ \alpha_1\beta_1 < 1. \tag{2.3.40}$$

首先,我们证明在条件(2.3.40)下,动态系统(2.3.39)是几何速度稳定的. 记

$$T(x_{t-1}) = \varphi(x_{t-1}) + H(x_{t-1}),$$

其中

$$\varphi(x_{t-1}) = \alpha_1 I(x_{t-1} \geqslant 0)x_{t-1} + \beta_1 I(x_{t-1} < 0)x_{t-1},$$

$$H(x_{t-1}) = \alpha_0 I(x_{t-1} \geqslant 0) + \beta_0 I(x_{t-1} < 0),$$

则动态系统(2.3.39)可表示为:

$$\begin{cases} x_t = T(x_{t-1}) + H(x_{t-1}); \\ x_0 \in R^1. \end{cases} \tag{2.3.41}$$

(i) 当 $|\alpha_1|<1, |\beta_1|<1$ 时,易见

$$|T(x)| = |\varphi(x) + H(x)| \leqslant \max\{|\alpha_1|, |\beta_1|\}|x| + (|\alpha_0| + |\beta_0|)$$
$$= \rho|x| + c,$$

其中 $\rho = \max\{|\alpha_1|, |\beta_1|\},\ c = |\alpha_0| + |\beta_0|.$

故由定理 2.3.12 知,此时动态系统(2.3.39)在有界区间

$$L = \{x : |x| \leqslant c/(1-\rho)\}$$

处为几何速度稳定的.

(ii) 当 $0<\alpha_1<1, \beta_1<-1$ 时,必存在正整数 q,使得

$$\alpha_1^{q-1}|\beta_1| = \rho < 1. \tag{2.3.42}$$

当 $x_0>\alpha_1^{-q}|\alpha_0|/(1-\alpha_1)$ 时,由(2.3.41)式有

$$x_1 = T(x_0) = \alpha_1 x_0 + \alpha_0$$

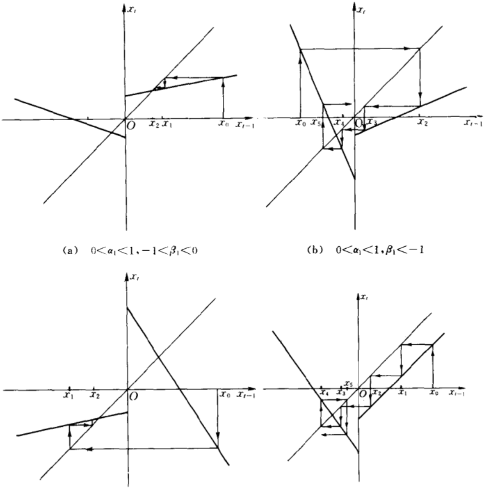
$$\geqslant \alpha_1^{-q+1}|\alpha_0|/(1-\alpha_1)+\alpha_0>0,$$

$$x_2=T(x_1)=\alpha_1 x_1+\alpha_0=\alpha_1^2 x_0+\alpha_0\alpha_1+\alpha_0$$

$$\geqslant \alpha_1^{-q+2}|\alpha_0|/(1-\alpha_1)-|\alpha_0|/(1-\alpha_1)>0.$$

依此类推，可得

$$x_q=T(x_{q-1})=\alpha_1^q x_0+\alpha_0(1+\alpha_1+\cdots+\alpha_1^{q-1}). \quad (2.3.43)$$



(a)　$0<\alpha_1<1, -1<\beta_1<0$　　(b)　$0<\alpha_1<1, \beta_1<-1$

(c)　$\alpha_1<-1, 0<\beta_1<1$　　(d)　$\alpha_1=1, \alpha_0<0, \beta_1<-1, \beta_0<0$

图 2.3.4　动态系统(2.3.39)四种不同情况的图形以及迭代示意图

当 $x_0<-\alpha_1^{-q}|\alpha_0|/(1-\alpha_1)-|\beta_0|$时，由(2.3.41)式有

$$x_1=T(x_0)=\beta_1 x_0+\beta_0$$

$$>|\beta_1|\alpha_1^{-q}|\alpha_0|/(1-\alpha_1)+|\beta_0\beta_1|+\beta_0$$

$$>\alpha_1^{-q}|\alpha_0|/(1-\alpha_1)>0,$$

$$
\begin{aligned}
x_2 &= T(x_1) = \alpha_1 x_1 + \alpha_0 > 0, \\
x_3 &= T(x_2) = \alpha_1 x_2 + \alpha_0 > 0, \\
&\cdots\cdots \\
x_q &= T(x_{q-1}) = \alpha_1^{q-1} x_1 + \alpha_0(1 + \alpha_1 + \cdots + \alpha_1^{q-2}) \\
&= \alpha_1^{q-1}\beta_1 x_0 + \alpha_1^{q-1}\beta_0 + \alpha_0(1 + \alpha_1 + \cdots + \alpha_1^{q-2}).
\end{aligned}
\tag{2.3.44}
$$

联合(2.3.43)和(2.3.44)两式，得

$$
\begin{aligned}
|T_q(x)| &\leqslant \max\{\alpha_1^q, \alpha_1^{q-1}|\beta_1|\}|x| + c \\
&= \rho|x| + c,
\end{aligned}
\tag{2.3.45}
$$

式中 $0<\rho<1$，c 为适当的常数. 故由定理 2.3.16 知，当 $0<\alpha_1<1$，$\beta_1<-1$时，动态系统(2.3.39)在有界区间

$$L = \{x: |x| \leqslant c/(1-\rho)\}$$

处为几何速度稳定的.

(iii) 当 $0<\beta_1<1$，$\alpha_1<-1$ 时，经变换 $\tilde{x}_t = -x_t$ 后，(2.3.39)式可化为

$$
\tilde{x}_t = \begin{cases} -\beta_0 + \beta_1 \tilde{x}_{t-1}, & \text{当 } \tilde{x}_{t-1} > 0; \\ -\alpha_0 + \alpha_1 \tilde{x}_{t-1}, & \text{当 } \tilde{x}_{t-1} \leqslant 0. \end{cases}
$$

于是，类似于(ii)的讨论知，存在一有界区间，使得动态系统(2.3.39)在其处是几何速度稳定的.

类似地可证明，在(2.3.40)的其他情况下，动态系统(2.3.39)也是几何速度稳定的. 详细的讨论留给读者.

下面，我们证明条件(2.3.40)也是动态系统(2.3.39)为几何速度稳定的必要条件. 为证此事实，分以下不同情况进行讨论.

(i) 若 $\alpha_1>1$，则必存在 $a>0$，使得 $(\alpha_1-1)a+\alpha_0>0$，从而，当 $x_0>a$ 时，由(2.3.39)式得

$$
\begin{aligned}
x_1 &= T(x_0) = \alpha_1 x_0 + \alpha_0 = x_0 + (\alpha_1 - 1)x_0 + \alpha_0 > x_0, \\
x_2 &= T(x_1) = \alpha_1 x_1 + \alpha_0 = x_1 + (\alpha_1 - 1)x_1 + \alpha_0 > x_1, \\
&\cdots\cdots
\end{aligned}
$$

依此类推，有

$$x_t = T_t(x_0) \to \infty, \quad \text{当 } t \to \infty \text{ 时}.$$

上式表明动态系统(2.3.39)是发散的，从而不可能为稳定的，更不可能为几何速度稳定的.

(ii) 若 $\beta_1>1$，则用类似于(i)的方法可证明，动态系统(2.3.38)不可

能是几何速度稳定的.

(iii) 若 $\alpha_1=1$,且 $\alpha_0>0$,则对任意 $x_0>0$,都有

$$T_{t-1}(x_0)<T_t(x_0)\to\infty,\ \text{当}\ t\to\infty\ \text{时}.$$

若 $\alpha_1=1$,而 $\alpha_0=0$,则动态系统(2.3.39)的吸引集 $L\supset(0,\infty)$. 由性质 2.3.4 知,动态系统(2.3.39)不是稳定的.

若 $\alpha_1=1$ 且 $\alpha_0<0$,按照对动态系统(2.3.10)的讨论知,动态系统(2.3.39)不是几何速度稳定的.

(iv) 对 $\beta_1=1$,而 $\beta_0>0$,$\beta_0=0$ 或 $\beta_0<0$ 的情况,依(iii)的证法可知,动态系统(2.3.39)不是几何速度稳定的.

(v) 若 $\alpha_1<1$,$\beta_1<1$,但 $\alpha_1\beta_1\geqslant1$,则必有 $\alpha_1<0$ 且 $\beta_1<0$,或者 $\alpha_1>0$,$\beta_1>0$. 不妨只讨论 $\alpha_1>0$ 且 $\beta_1>0$ 的情况. 前一情况可类似地讨论. 由于 $\alpha_1\beta_1\geqslant1$ 及 $\alpha_1>0$,$\beta_1>0$,则必有 $\alpha_1\geqslant1$,或者 $\beta_1\geqslant1$。由上述已讨论的情形知,不论 $\alpha_1\geqslant1$,还是 $\beta_1\geqslant1$,动态系统(2.3.39)都不是几何速度稳定的.

综上所述,条件(2.3.40)是动态系统(2.3.39)为几何速度稳定的充要条件.

例 2.3.23 继续例 2.2.12 的讨论,在此讨论动态系统(2.2.24)的稳定性. 如 §2.2 中所述,动态系统(2.2.24)的吸引集为 $L=\{(c,c)^{\tau}:0\leqslant c\leqslant1\}$,而且每个元素 $(c,c)^{\tau}$ 都是极限点. 它的吸引域为某个椭圆轨道在 D 中的部分(见图 2.2.4). 由此可见,它们都不是定义 2.3.6 所定义的稳定不动点.

我们还要指出一点:当动态系统(2.2.2)中的 D 为有界闭集时,如 §2.2 中的例 2.2.12 中的 $D=[0,1]\times[0,1]$,无论 $\boldsymbol{x}_0\in D$ 为何值,动态系统(2.2.2)的迭代序列是有界序列. 从而必在 D 中有极限点. 所以在此情形下,动态系统(2.2.2)的吸引集不会是空集. 但是当 D 为无界集时,例如动态系统(2.1.1)中 $D=R^1$,则动态系统的吸引集可能是空集. 比如动态系统(2.1.1)的情况(iii)就是如此. 如果在动态系统(2.2.2)中不仅 D 为有界闭集,而且 T 还是连续的,则吸引集还有更多的性质,读者可参阅文献[106].

例 2.3.24 考虑如下分段线性动态系统:

$$x_t=\begin{cases}1.8x_{t-1}-0.9x_{t-2}, & \text{当}\ x_{t-2}>0;\\ -0.9x_{t-1}, & \text{当}\ x_{t-2}\leqslant0,\end{cases}\qquad(2.3.46)$$

$$x_0\in R^1.$$

取 $x_0=x_{-1}=a<0$,于是有

$$x_1 = -0.9x_0 = -0.9a > 0,$$
$$x_2 = -0.9x_1 = (-0.9)^2 a < 0,$$
$$x_3 = 1.8x_2 - 0.9x_1 = 2(0.9)^3 a + (0.9)^2 a < 0,$$
$$x_4 = -0.9x_3 = -2.8(0.9)^3 a > 0,$$
$$x_5 = -0.9x_4 = 2.8(0.9)^4 a < 0,$$
$$x_6 = 1.8x_5 - 0.9x_4$$
$$= 1.8(2.8)(0.9)^4 a + (2.8)(0.9)^4 a < 0.$$

依此类推可知：

$$x_{3k+1} > 0,\ x_{3k+2} < 0,\ x_{3k+3} < 0.$$

并且当 $t \to \infty$ 时，

$$|x_t| \to \infty.$$

从而动态系统(2.3.46)有发散点，故不可能稳定.

值得指出的是：动态系统(2.3.46)是由两个线性系统组成的分段线性系统，而且每个线性系统都在零点处稳定. 但是由它们组成的分段线性系统(2.3.46)却不是稳定的.

作为本章的结束，我们再举一例，初步阐释非线性动态系统与分形几何的关系，在 § 8.3 中，我们要进一步讨论非线性时间序列与分形维数的联系.

例 2.3.25 考虑如下二维动态系统

$$\begin{cases} \boldsymbol{x}_t = \boldsymbol{T}(\boldsymbol{x}_{t-1}) \quad (t \geqslant 1); \\ \boldsymbol{x}_0 \in R^2, \end{cases} \tag{2.3.47}$$

其中 $\boldsymbol{x}_t = (x_{1,t}, x_{2,t})^\tau$,

$$\boldsymbol{T}(\boldsymbol{x}_{t-1}) = \begin{pmatrix} \varphi_1(\boldsymbol{x}_{t-1}) \\ \varphi_2(\boldsymbol{x}_{t-1}) \end{pmatrix} = \begin{pmatrix} x_{1,t-1}^2 - x_{2,t-1}^2 + c_1 \\ 2x_{1,t-1} - x_{2,t-1} + c_2 \end{pmatrix}.$$

由(2.3.47)式可知

$$x_{1,t} = x_{1,t-1}^2 - x_{2,t-1}^2 + c_1, \tag{2.3.48}$$

$$x_{2,t} = 2x_{1,t-1} - x_{2,t-1} + c_2. \tag{2.3.49}$$

以上两式中的 c_1 和 c_2 是常数参数.

如果记 $z_t = x_{1,t} + ix_{2,t}$, $c = c_1 + ic_2$ 为复数，则由(2.3.48)和(2.3.49)式可得如下等价的复值动态系统

$$\begin{cases} z_t = z_{t-1}^2 + c \equiv Q(z_{t-1}) \quad (t \geqslant 1); \\ z_0 \in Z, \end{cases} \tag{2.3.50}$$

式中 $Q(z)=z^2+c$，Z 为复平面．对比(2.3.47)式和(2.3.50)式可知，后者仅是前者的复数表示，即将(2.3.47)式中 $\boldsymbol{x}_t$ 的第一分量当作实数部分，第二分量视为虚数部分，就可得到复值动态系统(2.3.50)，它与(2.3.47)是完全等价的．但是研究复值动态系统(2.3.50)的吸引集和吸引域会更有意义，也更方便．

首先考虑 $c=0$ 的情况，即(2.3.50)式为

$$\begin{cases} z_t = z_{t-1}^2 \quad (t \geqslant 1); \\ z_0 \in Z. \end{cases} \tag{2.3.51}$$

易见，当 $t\to\infty$ 时：

(i) 若 $|z_0|<1$，则 $z_t\to 0$；

(ii) 若 $|z_0|>1$，则 $|z_t|\to\infty$；

(iii) 若 $|z_0|=1$，则 $|z_t|=1$．

以上三条表明了复平面的原点 $z_0=0$ 是动态系统(2.3.51)的极限点，也是 $Q(z)=z^2$ 的不动点．这里所称的极限点以及以下的极限环等，既可理解为相应的动态系统(2.3.47)的极限点和极限环，也可理解为 § 2.2 中极限点和极限环在复值动态系统的推广．根据上面的性质(i)还可知，$z_0=0$的吸引域为单位圆的全体内点，即$\{z:|z|<1\}$．再根据性质(ii)知，若将无穷远点视为另一个吸引点，则$\{z:|z|>1\}$是其吸引域．最后根据(iii)知，若以单位圆周上的点 $z_0=e^{i2\pi\theta_0}$为出发点，则由(2.3.51)式迭代的序列$\{z_t\}$都在单位圆周上．若记 $z_t=e^{i2\pi\theta_t}$，并限定 $0\leqslant\theta_t<1$，由(2.3.51)式可知，θ_t 满足如下的迭代关系式，即

$$\begin{cases} \theta_t = 2\theta_{t-1}\mathrm{mod}(1) \quad (t \geqslant 1); \\ \theta_0 \in [0,1). \end{cases} \tag{2.3.52}$$

动态系统(2.3.52)与(2.2.13)式相同，关于对它的吸引集的讨论可见例 2.2.9.

如果只将有穷极限作为极限点，那么动态系统(2.3.51)的吸引集为

$$L_{L0} = \{z:|z| = 1\} \cup \{0\},$$

式中足标“0”表示(2.3.50)式中的 $c=0$．又依上述的性质(i)～(iii)知，L_0 的吸引域为

$$A_T(L_0) = \{z:|z| \leqslant 1\}. \tag{2.3.53}$$

下面简单讨论动态系统(2.3.50)中 $c\neq 0$ 的情况．

(i) 当 $c\neq 0$ 时，变换 $Q(z)=z^2+c$ 有任意长度的周期解．易见 $Q(z)=z$ 总在 Z 中存在解，从而必有周期为 1 的解．为了证明 $Q(z)$有长度为 p

的周期解，只要证明 $Q(z)$ 经 p 次迭代后的函数 $Q_p(z)=z$ 在 Z 中有解即可，而这是显然的.

(ii) 依上面的性质又知，$Q(z)$ 的吸引集为非空，从而 $Q(z)=z^2+c$ 的吸引集的吸引域也非空.

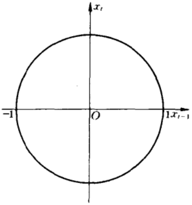


图 2.3.5 $Q(z)=z^2$ 的吸引域

(iii) 对于固定的 $c\neq 0$，只要(2.3.50)式中的初值 z_0 的模足够大，当 $t\to\infty$ 时，必有 $|z_t|\to\infty$. 为说明这一事实，记 $r_t=|z_t|$，$\rho=|c|$，$\delta>0$ 为任意固定的常数. 当初值满足

$$|z_0|^2-|c|=r_0^2-\rho>r_0+\delta \tag{2.3.54}$$

并注意到 $r_0\geqslant 0$ 时，上式与下式等价，即

$$|z_0|=r_0>\frac{1}{2}\{1+(1+4\rho+4\delta)^{1/2}\}, \tag{2.3.55}$$

则 $z_1=Q(z_0)=z_0^2+c$ 必满足

$$\begin{aligned}r_1=|z_1|&=|z_0^2+c|\geqslant|z_0|^2-|c|=r_0^2-\rho>r_0+\delta\\&>\frac{1}{2}\{1+(1+4\rho+4\delta)^{1/2}\}.\end{aligned}\tag{2.3.56}$$

此式表明了若视 z_1 为(2.3.50)式的初始值时，它也满足(2.3.54)和(2.3.55)式，从而类似于(2.3.56)式的推导，又有

$$\begin{aligned}r_2&=|z_2|=|z_1|^2-|c|\\&=r_1^2-\rho>r_1+\delta>\frac{1}{2}\{1+(1+4\rho+4\delta)^{1/2}\}.\end{aligned}$$

于是用归纳法可得

$$\begin{aligned}r_t&=|z_t|=|z_{t-1}|^2-|c|=r_{t-1}^2-\rho>r_{t-1}+\delta>r_{t-1}\\&>\frac{1}{2}\{1+(1+4\rho+4\delta)^{1/2}\}.\end{aligned}$$

利用上式中的 $r_t>r_{t-1}+\delta$，又可得

$$r_t=|z_t|>r_{t-1}+\delta>\cdots>r_0+t\delta\to\infty,\ \text{当}\ t\to\infty\ \text{时}.$$

(iv) 根据上述结果又知 $Q(z)=z^2+c$ 的吸引集是有界集，不妨记为 L_c. 又由于对任何足够大的 $r_0=|z_0|$ 而言，(iii)的结论都成立，所以，L_c 的吸引域也是有界集.

除了以上较简单的结论外，要了解 $Q(z)=z^2+c\,(c\neq 0)$ 有哪些吸引

元？各自的吸引域是什么？则是非常复杂的事。这里仅举一两个例子来说明其复杂程度.

当 $c=-0.12375+0.56508i$ 时，$z^2+c=z$ 有两个解，它们是

$$z_{01}=1.2249995-0.3871049i;$$

$$z_{02}=0.2249995+0.3871049i.$$

回忆(2.3.47)和(2.3.50)的等价关系，可知动态系统(2.3.47)有如下两个不动点：

$$\boldsymbol{x}_{01}=\begin{pmatrix}1.2249995\\-0.3871049\end{pmatrix},\quad \boldsymbol{x}_{02}=\begin{pmatrix}0.2249995\\0.3871049\end{pmatrix}.$$

求动态系统(2.3.47)中变换 $\boldsymbol{T}$ 的偏导数，得

$$\frac{\partial \boldsymbol{T}}{\partial \boldsymbol{x}}=2\begin{bmatrix}x_1 & x_2\\-x_2 & x_1\end{bmatrix}.$$

不难验证，$\partial \boldsymbol{T}/\partial \boldsymbol{x}|_{x=x_{02}}$ 的特征值的模都小于 1，而对 $\partial \boldsymbol{T}/\partial \boldsymbol{x}|_{x=x_{01}}$，却并非如此. 由此可知，$\boldsymbol{x}_{02}$ 是动态系统(2.3.47)的稳定不动点；而 $\boldsymbol{x}_{01}$ 却不是稳定的. 从而在动态系统(2.3.50)中，z_{02} 是稳定的不动点，而 z_{01} 则不是稳定的. 此时，z_{02} 的吸引域是包含 z_{02} 点的一个连通集，如图 2.3.6 所示.

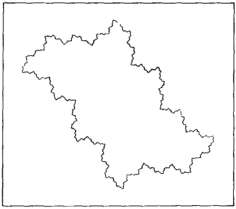


图 2.3.6 $Q(z)=z^2-0.12375+0.56508i$ 的吸引域

从连通性看，这与 $c=0$ 时的情况相似. 在 $c=0$ 时，$z_0=0$ 是稳定不动点，其吸引域为 $\{z: |z|<1\}$ 也是连通集。但是它们又有着本质性的差别. 吸引域 $\{z: |z|<1\}$ 的边界为 $\{z: |z|=1\}$（见图 2.3.5），其周长为 2π. 而 z_{02} 的吸引域（见图 2.3.6）的边界长度为无穷大. 当细看图 2.3.6 的边界时，若将它放在放大镜下看，其边界的曲折形状与原来一样. 这样的集合

属于分形几何学所研究的内容. 图 2.3.6 还告诉我们,变换 $Q(z)=z^2+c$ 的其他吸引元都落在曲折的边界线上,而边界的外部诸点被 $Q(z)$ 反复迭代而向无穷远发散.

对于某些 c 值而言,$Q(z)=z^2+c$ 的吸引集 L_c 的吸引域的边界可以不是连通集. 如图 2.3.7 所示的某个 $Q(z)$ 的吸引域,其边界并不连接.

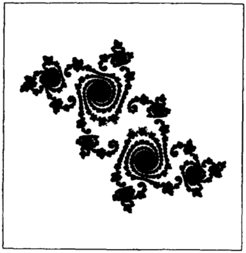


图 2.3.7 $Q(z)=z^2-0.194+0.6557i$ 的吸引域

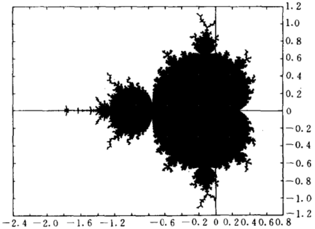


图 2.3.8 变换 $Q(z)=z^2+c$ 的 Mandelbrot 集

图中复数 c 平面的范围是 $-2.24<\mathrm{Re}\,c<0.8$, $-1.2<\mathrm{Im}\,c<1.2$.

此图反映不同参数值 c 的各种类型边界之间的有序性

如果我们将动态系统(2.3.50)中的参数 c 分为两类:一类是使 $Q(z)$

的吸引域的边界为连通集；另一类是使 $Q(z)$ 的吸引域的边界为不连通的. 在复平面 Z 上，把第一类中的 c 构成的集合着黑色，其补集为构成第二类 c 值的集合，如图 2.3.8 所示. 第一类中的 c 构成的集合称为 Mandelbrot 集（见文献[87]）.

通过以上对动态系统(2.3.50)亦即对动态系统(2.3.47)的简单讨论，可以看到如此简单的二维动态系统，它们的吸引集和吸引域的结构竟是如此复杂和丰富多彩. 揭示这些特征不仅要靠理论分析，更要借助于现代的计算技术. 这些内容都属于近代较活跃的分形几何学领域. 读者可参阅文献[87]和[120].

第 3 章

非线性时间序列模型

本章介绍近年来比较深入研究的、有些已被广泛应用的非线性时间序列模型. 当然，这些模型还不能详尽地描述内容丰富的非线性时间序列. 为了分清这些非线性时间序列模型在时间序列分析中的作用和局限性，我们在 §3.1 中先对时间序列的分类作一概述.

§3.1 时间序列分类概述

在第 1 章中，我们讨论了时间序列依平稳性的分类. 本节将对时间序列的分类作更广泛的讨论.

首先，时间序列$\{x_t, t\in T\}$可按 x_t 的取值分类. x_t 可为随机变量，随机向量，或更抽象空间的随机元. 本书主要讨论随机变量的情况. 虽然有时也涉及随机向量的情况，但其目的仅在于处理高阶自回归时间序列模型.

时间序列也可按时间的取值范围分类，例如 t 的取值范围 T 可以是所有正整数，这是“时间”的本来含义，也可推广到二维整数对或更高维的情况. 这时，“时间”实际上是空间. 有时将高维情况称为“时空”，其含义是：t 的第一分量表示“时间”，其他分量则表示“空间”. 总之，t 为高维时，“时间”序列实际上是场序列，或者是时空序列. 本书只讨论 t 为整数的情况. 当然，本书的理论、方法和技巧对研究其他的情况可能也有所借鉴.

时间序列还有其他分类，最重要的是平稳与非平稳；也可按其矩的性质进行分类，这主要是在理论分析中使用. 用高阶矩研究非线性时间序列已在 §1.3 中叙述过了.

对于平稳序列，有线性与非线性的分类，这是本书讨论的重点. 时间序列的“线性性”的定义不止一种，随着对非线性时间序列的深入研究，关于线性性的不同涵义的界定是不容忽视的. 下面我们通过例子来说明这一点.

考虑如下的时间序列：

$$x_t = \varepsilon_t + \sum_{k=1}^{\infty} \varepsilon_{t-k} \prod_{i=0}^{k-1} (\alpha + \beta\varepsilon_{t-i}), \tag{3.1.1}$$

其中$\{\varepsilon_t\}$为白噪声序列，ε_t服从正态$N(0,1)$分布，参数α和β满足$\alpha^2+\beta^2<1$. 易见，(3.1.1)式定义的随机级数在均方意义下是收敛的，且

$$Ex_t = 0,$$
$$Ex_t^2 = (1-\alpha^2-\beta^2)^{-1} < \infty.$$

容易验证时间序列(3.1.1)是平稳的，但是按定义1.2.1知，它不是线性的. 另外，用直接验证可知，x_t满足如下随机模型

$$x_t = \alpha x_{t-1} + \beta x_{t-1}\varepsilon_t + \varepsilon_t. \tag{3.1.2}$$

由后面§3.4中的讨论可知，模型(3.1.2)属于双线性模型，是一类非线性时间序列模型. 但是，由(3.1.1)和(3.1.2)式根据(1.4.1)式计算x_t的一步条件期望预报，可得到

$$\begin{aligned}\hat{x}_n(1) &= E\{x_{n+1} | x_n, x_{n-1}, \cdots\} \\ &= E\{\alpha x_n + \beta x_n \varepsilon_{n+1} + \varepsilon_{n+1} | x_n, x_{n-1}, \cdots\} \\ &= \alpha x_n.\end{aligned}$$

不难验证，对满足(1.4.2)式的$\hat{x}_n^*(1)$，有

$$\hat{x}_n^*(1) = \hat{x}_n(1) = \alpha x_n.$$

这就表明，从预报的观点来看，由(3.1.2)式确定的时间序列$\{x_t\}$是预报线性序列.

通过上述分析可知，在第1章中给出的线性序列的两种定义确有不同的涵义，应当加以区分，这与§1.4中的新息序列的假定的分类是有关的. 在以后有关模型结构和预报的章节中，我们要仔细讨论这些内容.

按一步预报性质，也可对时间序列进行分类，即按

$$\hat{x}_n(1) = E\{x_{n+1} | x_n, x_{n-1}, \cdots\} \equiv \varphi(x_n, x_{n-1}, \cdots) \tag{3.1.3}$$

中的预报函数φ进行分类.

当预报函数φ仅依赖于$x_n, x_{n-1}, \cdots, x_{n-p+1}$等有限个历史值，即

$$\begin{aligned}\hat{x}_n(1) &= E\{x_{n+1} | x_n, x_{n-1}, \cdots\} \\ &\equiv \varphi(x_n, x_{n-1}, \cdots, x_{n-p+1})\end{aligned} \tag{3.1.4}$$

时，则由(1.4.4)式知，新息序列按下式定义：

$$e_{n+1} = x_{n+1} - \hat{x}_n(1). \tag{3.1.5}$$

于是有

$$x_{n+1} = \varphi(x_n, x_{n-1}, \cdots, x_{n-p+1}) + e_{n+1}. \tag{3.1.6}$$

因此，如果时间序列$\{x_t\}$的预报函数φ仅依赖于p个历史值，即(3.1.4)

式成立,则称其为 p 阶自回归序列. 否则,称其为无穷阶自回归序列. 这时,对时间序列$\{x_t\}$的进一步的分类,要依赖于预报函数 φ 的形式和新息序列$\{e_t\}$的假定. 若 φ 被有限个参数所决定,则称(3.1.6)为参数模型;若 φ 是非参数型的任意函数,则称(3.1.6)为非参数模型. 而参数模型又有线性与非线性之分. 这些已在§1.3中有过讨论了. 下面从分类角度再来分析新息序列$\{e_t\}$的不同假定的作用.

如果$\{e_t\}$被假定为平稳鞅差序列,则是对模型(3.1.6)的一种较粗略的分类. 比较精细的分类需要对$\{e_t\}$作进一步的假定. 例如,最有用的一类假定是

$$e_t = \varepsilon_t S(x_{t-1}, \cdots, x_{t-p}), \tag{3.1.7}$$

其中$\{\varepsilon_t\}$为白噪声序列. 这就导致以下更精细的模型:

$$x_t = \varphi(x_{t-1}, \cdots, x_{t-p}) + \varepsilon_t S(x_{t-1}, \cdots, x_{t-p}). \tag{3.1.8}$$

值得指出的是,在(3.1.7)式中仍可假定$\{\varepsilon_t\}$为平稳鞅差序列. 这样,我们又称(3.1.7)式是一个较粗略的分类. 因为,如果假定$\{\varepsilon_t\}$是一个白噪声序列,这就意味着除了一维分布未知外,对$\{\varepsilon_t\}$的概率结构完全给定了. 所以关于时间序列模型的假定,当它仅依赖于一个白噪声序列时,才算是比较精细的模型. 从这一观点来看,当模型(3.1.6)中的$\{e_t\}$被假定为白噪声序列时,模型(3.1.6)才为精细的模型. 在$\{e_t\}$被假定为平稳鞅差序列时,除了引用(3.1.7)式的假定外,其他类型的假定尚未有讨论.

从上述讨论可见,在§1.4中以 $\hat{x}_n(1)=\hat{x}_n^*(1)$ 作为时间序列线性性的定义时,实际上并没有对序列的结构给出精确的界定. 而定义1.2.1所定义的线性性,不仅保证了其预报的线性性,而且它还是一白噪声序列的滑动和. 这就使得时间序列的结构的假定比较精细了. 由此不难理解,(3.1.1)式定义的时间序列在预报意义下为线性序列,而在定义1.2.1的意义下不一定是线性序列. 从此意义上说,定义1.2.1的线性序列是比较精细的分类.

在非线性时间序列中,除了按(3.1.6)式中的预报函数 φ 和新息序列$\{e_t\}$进行分类外,还有类似于线性序列中LAR、LMA和LARMA序列在非线性情况下的各种推广. 这些模型至少有两种来源,一种是形式地从线性模型推广到非线性;另一种是从实际中的微分方程附带噪声项所引出的随机差分方程,它们可能呈现出NLARMA模型的形式. 关于这些模型的分类、平稳性和遍历性的讨论是对相应的线性模型的推广. 需要指出的是,满足这些模型的时间序列,即使是平稳的,它们的预报函数 φ 也不

一定属于前述的有限阶的情形，多数是无穷阶的.

在这一节中关于时间序列的分类，仅仅涉及了本书以后章节中提到的非线性模型，只能算作对时间序列的初步分类.

§ 3.2　可加噪声模型

在 § 3.1 中，我们从预报角度对时间序列进行分类时，引出了非线性自回归模型(3.1.6). 如第 1 章中所述，从预报观点来看，模型(3.1.6)中的$\{e_t\}$只宜被假定为平稳鞅差序列，从而还不能对模型(3.1.6)作出较精确的界定，例如在$\{e_t\}$为平稳鞅差序列的假定下，模型(3.1.6)至少包括了形如(3.1.8)式的模型，其中函数 S 还可有多种类型的选择. 所以本节所论及的可加噪声模型，是指模型(3.1.6)中$\{e_t\}$为白噪声序列的情况. 此类可加噪声模型按对 φ 的不同假定，可分为参数、非参数和半参数三类模型.

一、非参数自回归模型

非参数自回归模型的一般形式为

$$x_t = \varphi(x_{t-1}, \cdots, x_{t-p}) + \varepsilon_t, \qquad (3.2.1)$$

其中 φ 为 R^p 到 R^1 的可测函数，$\{\varepsilon_t\}$为白噪声序列，满足条件(1.3.7)式，即

$$\begin{cases} E\varepsilon_t = 0,\ E\varepsilon_t^2 = \sigma^2 < \infty; \\ \varepsilon_t \text{ 与 } \{x_s,\ s < t\} \text{ 相互独立}. \end{cases} \qquad (3.2.2)$$

模型(3.2.1)还有如下两种特殊情况：

(一) 可加非线性自回归模型(Additive Nonlinear Autoregression)

可加非线性自回归模型定义为

$$x_t = c + f_1(x_{t-1}) + \cdots + f_p(x_{t-p}) + \varepsilon_t, \qquad (3.2.3)$$

其中 c 为常数，$f_i(i=1,\cdots,p)$为 p 个一元非参数型的未知函数，$\{\varepsilon_t\}$为白噪声序列，满足(3.2.2)式. 为了简便，把模型(3.2.3)记为 ANLAR(p)模型，p 称为模型(3.2.3)的阶数.

(二) 函数系数自回归模型(Functional Coefficient Autoregression)

函数系数自回归模型定义为

$$x_t = c + f_1(x_{t-d})x_{t-1} + \cdots + f_p(x_{t-d})x_{t-p} + \varepsilon_t, \qquad (3.2.4)$$

其中 c 为常数，$f_i(i=1,\cdots,p)$为 p 个一元非参数型的未知函数，$0<d\leqslant p$

为整数，称为*延迟参数*，$\{\varepsilon_t\}$为白噪声序列，满足(3.2.2)式. 为了简便，把模型(3.2.4)记为 FCAR(p,d)，p 称为模型(3.2.4)的*阶数*.

模型(3.2.4)可以很好地解释为一维局部线性近似. 具体地说，如果函数 $f_1,\cdots,f_p$ 都较光滑，则在 x_{t-d}的邻域内可用常数值近似. 在这情况，当 x_{t-d}属于 $x_{t-d}=a$ 的邻域时，(3.2.4)式可近似地表为

$$x_t = c + f_1(a)x_{t-1} + \cdots + f_p(a)x_{t-p} + \varepsilon_t. \qquad (3.2.5)$$

这一想法与以下的(3.2.7)式所述的门限自回归模型很近似，并且这两类模型的拟合方法也有相似之处.

模型(3.2.1)中的函数 φ 未被限定类型，为了理论研究或统计分析的方便，可对 φ 的光滑性作某些假定，例如设 φ 为分片连续函数，连续函数，或者连续可微函数等. 为了保证模型(3.2.1)有平稳解和遍历性，还需要求函数 φ 满足适当的约束条件. 此外，模型(3.2.1)的结构还与白噪声 ε_t 的一维分布有关. 这些内容都将在第 4 章中讨论. 在此还需指出，当 φ 为非线性函数时，即使$\{\varepsilon_t\}$为正态白噪声序列，而由(3.2.1)式确定的随机序列也不一定是正态序列.

满足模型(3.2.1)的时间序列$\{x_t\}$，其被观测到的样本序列在一定意义下表现着模型(3.2.1)的特征. 由于函数 φ 和 ε_t 的分布是未知的，根据模型(3.2.1)的样本序列透视模型的特征，是对该模型进行统计分析的重要依据. 这些内容将在第 6 章进行讨论，在此先举一例来说明由(3.2.1)式产生的样本序列的特征.

例 3.2.1 考虑如下一阶非线性模型

$$x_t = \frac{\alpha x_{t-1}}{1 + x_{t-1}^2} + \varepsilon_t, \qquad (3.2.6)$$

其中 $0<\alpha<2$，$\{\varepsilon_t\}$为正态白噪声序列，且满足(3.2.2)式. 由(3.2.6)式可产生$\{x_t\}$的模拟样本序列(不妨取 $x_0=0$)，图 3.2.1 给出了样本序列随时间变化的图形和直方图. 图 3.2.2 给出了(x_{t-1},x_t)的图形. 当我们得到模型(3.2.6)的样本序列时，如果不知道 $\varphi(x)=\alpha x/(1+x^2)$和$\{\varepsilon_t\}$服从正态分布等真实信息，那末通过图 3.2.1 和图 3.2.2 去分析 φ 的形状特征是有益的. 当然，对于高阶非线性模型，利用上面的绘图方法去观察分析序列的特征会碰到困难. 处理高阶自回归序列的一些近似手段将在以后讨论.

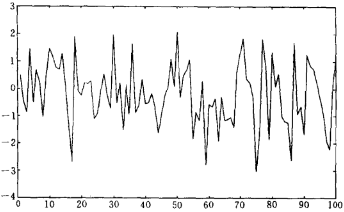
二、参数型自回归模型

当自回归模型(3.2.1)中的函数 φ 属于指定的函数类，但含有有限个

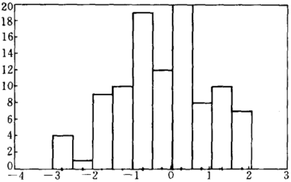
未知参数时，则称其为**参数型自回归模型**. 它的一般形式为

$$x_t = \varphi(x_{t-1}, \cdots, x_{t-p}; \boldsymbol{\theta}) + \varepsilon_t, \tag{3.2.7}$$

其中 $\boldsymbol{\theta}=(\theta_1,\cdots,\theta_s)^\tau \in \Theta$ 是未知参数，其维数 s 可与模型(3.2.7)的自回归阶数不同，$\{\varepsilon_t\}$ 为白噪声序列，满足(3.2.2)式.



(a)　样本序列随时间变化



(b)　样本序列的直方图

图　3.2.1

按照统计学中参数模型的概念，在(3.2.7)式中，关于 ε_t 的分布也应假定为由有限个参数所决定. 这样，模型(3.2.7)在$\{\varepsilon_t\}$为白噪声序列的总前提下，除去 φ 属于某指定的函数类之外，ε_t 的分布也属于某指定的分布类，且被有限参数所决定，这样模型(3.2.7)的全部结构仅被有限个参数所唯一确定. 例如，假定模型(3.2.7)中的 ε_t 服从正态 $N(0,\sigma^2)$分布，模型(3.2.7)的概率结构就被参数$(\boldsymbol{\theta}^\tau,\sigma^2)$所唯一确定. 但是，在时间序列分析中，人们关心的是模型(3.2.7)中函数 φ 的估计问题，而对 ε_t 的分布的

估计问题较少关心. 所以只假定 φ 为有限参数型时也称(3.2.7)式为参数型模型.

如果 φ 是 $x_{t-1},\cdots,x_{t-p}$ 的线性函数,即模型为线性自回归模型时,(3.2.7)式一定是参数型的. 参数 θ 的维数仍可不同于阶数 p. 例如疏系数 LAR(12)模型(见文献[2])

$$x_t = \theta_1 x_{t-1} + \theta_2 x_{t-12} + \varepsilon_t,$$

其中 $p=12$, $s=2$, $\theta=(\theta_1,\theta_2)^\tau$.

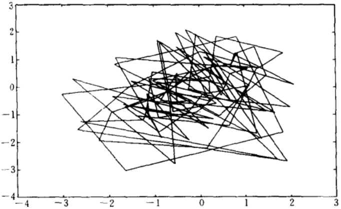


图 3.2.2 样本序列的 (x_t,x_{t-1}) 的图形

当 φ 被指定为不同函数类时,(3.2.7)就表现为不同的非线性自回归模型. 下面介绍几类研究得较多、且应用较为广泛的非线性自回归模型.

(一) 分片线性自回归模型

将空间 R^p 分割成 l 个互不相交的子集,使得 φ 在每个子集中为线性函数,这样的函数 φ 称为**分片线性函数**. 由分片线性函数 φ 生成的时间序列模型(3.2.7)称为**分片线性自回归模型**,由下式定义:

$$x_t = \sum_{j=1}^{l}\{\varphi_{j0} + \sum_{k=1}^{p}\varphi_{jk}x_{t-k}\}I(\boldsymbol{x}_{t-1} \in R_j) + \varepsilon_t, \qquad (3.2.8)$$

其中 $\boldsymbol{x}_t=(x_t,x_{t-1},\cdots,x_{t-p+1})^\tau$, $I(\cdot)$ 为示性函数, $R_j(j=1,\cdots,l)$ 为 R^p 的互不相交的子集,且 $R^p=R_1\cup R_2\cup\cdots\cup R_p$.

易见,模型(3.2.8)的参数向量 θ 由 $\{\varphi_{j0},\varphi_{jk},j=1,\cdots,l;k=1,\cdots,p\}$ 组成,其维数为 $s=l(p+1)$. 此模型的明显缺点是参数个数过多, R_j $(j=1,\cdots,l)$ 被确定也比较困难.

特别地,当模型(3.2.8)中的分割具有如下形式:

$$R_j = \{\boldsymbol{x} = (x_1,\cdots,x_p)^\tau : r_j \leqslant x_d < r_{j+1}, j = 1,\cdots,l\},$$

其中 $l \leqslant d \leqslant p$ 为某个整数，便得到应用最为广泛的一类参数型非线性自回归模型——门限自回归模型(Threshold Autoregression)，其定义为

$$x_t = \sum_{j=1}^{l} \{\varphi_{j0} + \sum_{k=1}^{p} \varphi_{jk} x_{t-k}\} I(r_j \leqslant x_{t-d} < r_{j+1}) + \varepsilon_t, \quad (3.2.9)$$

其中

$$-\infty = r_1 < r_2 < \cdots < r_l < r_{l+1} = +\infty,$$

$\{\varepsilon_t\}$ 为白噪声序列，满足(3.2.2)式，正整数 d 称为延迟参数，$r_2, \cdots, r_l$ 称为门限参数，p 为模型的阶数. 模型(3.2.9)简记为 TAR 模型.

例 3.2.2　考虑如下的 TAR 模型：

$$x_t = \begin{cases} 1.0 + 0.6x_{t-1} + \varepsilon_t, & \text{当 } x_{t-1} \geqslant 0; \\ -1.0 + 0.4x_{t-1} + \varepsilon_t, & \text{当 } x_{t-1} < 0, \end{cases} \quad (3.2.10)$$

其中 $\{\varepsilon_t\}$ 为白噪声序列，且 $\varepsilon_t \sim N(0,1)$.

图 3.2.3 是 TAR 模型(3.2.10)的样本数据图，而图 3.2.4 则是 LAR(1)模型

$$x_t = 0.6x_{t-1} + \varepsilon_t \quad (3.2.11)$$

的样本数据图，其中 $\varepsilon_t \sim N(0,1)$.

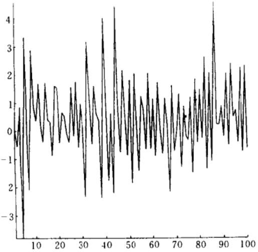


图 3.2.3　TAR 模型(3.2.10)的样本数据

分片线性自回归模型或者门限自回归模型，其回归函数有跳跃变化. 这种特征恰好刻划了自然界的突变现象. 例如在经济领域，许多指标受到多种因素的影响，使某些观测序列呈跳跃变化. 在水文、气象等领域中也有类似现象. 这就使得这类模型具有了一定的实际应用背景.

（二）拟线性自回归模型

拟线性自回归模型被定义为

$$x_t = \varphi_0 + \varphi_1 f_1(x_{t-1},\cdots,x_{t-p}) + \cdots + \varphi_s f_s(x_{t-1},\cdots,x_{t-p}) + \varepsilon_t, \tag{3.2.12}$$

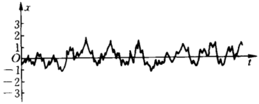


图 3.2.4 LAR(1)模型(3.2.11)的样本数据

其中 $f_i(i=1,\cdots,s)$ 是 s 个已知的 R^p 到 R^1 的可测函数，$\{\varepsilon_t\}$ 为白噪声序列，满足(3.2.2)式，$\varphi_i(i=1,\cdots,s)$ 是实参数.

由于模型(3.2.12)中的 f_i 可以是非线性函数，所以模型(3.2.12)属于非线性自回归模型. 但是模型(3.2.12)中的参数是以线性形式出现的，故称为拟线性自回归模型. 从形式上看，(3.2.12)与线性回归模型

$$y_t = \varphi_0 + \varphi_1 x_{1t} + \cdots + \varphi_s x_{st} + \varepsilon_t \tag{3.2.13}$$

是相同的. 所以用于回归模型(3.2.13)的统计方法，例如参数的最小二乘估计方法，完全适用于模型(3.2.12). 但是在模型(3.2.12)中，$y_t = x_t, x_{it} = f_i(x_{t-1},\cdots,x_{t-p})$ 是 x_t 的 p 步延迟函数，这使得模型(3.2.12)是否有平稳遍历解的问题被提出来. 而回归模型(3.2.13)中的 x_{it} 往往是非随机量测，或者 y_t 与 x_{it} 是不同随机变量的独立量测. 所以(3.2.13)式中的序列 $\{y_t\}$ 没有平稳解的问题.

如果分片线性自回归模型(3.2.8)中的分割区域 $R_1,\cdots,R_l$ 是完全给定的，不难看出，模型(3.2.8)也具有(3.2.12)的形式，即为拟线性自回归模型. 以 $l=p=2$ 为例，记 $\boldsymbol{x}_{t-1}=(x_{t-1},x_{t-2})^{\tau}$. 其模型为

$$\begin{aligned} x_t = &\{\varphi_{10} + \varphi_{11}x_{t-1} + \varphi_{12}x_{t-2}\}I(\boldsymbol{x}_{t-1} \in R_1) \\ &+ \{\varphi_{20} + \varphi_{21}x_{t-1} + \varphi_{22}x_{t-2}\}I(\boldsymbol{x}_{t-1} \in R_2) + \varepsilon_t \\ = &\varphi_{10} + (\varphi_{20} - \varphi_{10})I(\boldsymbol{x}_{t-1} \in R_2) + \varphi_{11}x_{t-1}I(\boldsymbol{x}_{t-1} \in R_1) \\ &+ \varphi_{12}x_{t-2}I(\boldsymbol{x}_{t-1} \in R_1) + \varphi_{21}x_{t-1}I(\boldsymbol{x}_{t-1} \in R_2) \\ &+ \varphi_{22}x_{t-2}I(\boldsymbol{x}_{t-1} \in R_2) + \varepsilon_t. \end{aligned}$$

将上式和(3.2.12)式相比较，不难看出：

$$f_1(x_{t-1},x_{t-2}) = I((x_{t-1},x_{t-2})^{\tau} \in R_2),$$

$$f_2(x_{t-1},x_{t-2}) = I((x_{t-1},x_{t-2})^\tau \in R_1)x_{t-1},$$
$$f_3(x_{t-1},x_{t-2}) = I((x_{t-1},x_{t-2})^\tau \in R_1)x_{t-2},$$
$$f_4(x_{t-1},x_{t-2}) = I((x_{t-1},x_{t-2})^\tau \in R_2)x_{t-1},$$
$$f_5(x_{t-1},x_{t-2}) = I((x_{t-1},x_{t-2})^\tau \in R_2)x_{t-2}.$$

由于上式中区域 R_1 和 R_2 是给定的,互不相交,且 $R^2=R_1\cup R_2$,故示性函数 $I((x_{t-1},x_{t-2})^\tau\in R_1)$和 $I((x_{t-1},x_{t-2})^\tau\in R_2)$也是给定的.因此上述的诸函数 f_i 是已知函数.从而上述二阶分片线性自回归模型也是拟线性自回归模型.

当 $p=1$ 时,(3.2.12)式有一种重要的特殊形式,即

$$x_t = \varphi_0 + \varphi_1 x_{t-1} + \varphi_2 x_{t-1}^2 + \varphi_3 x_{t-1}^3 + \sum_{k=4}^{l}\varphi_k(x_{t-1} - r_{k-3})^3 I(x_{t-1} > r_{k-3}) + \varepsilon_t, \quad (3.2.14)$$

式中自回归函数恰好是一个三次样条函数,即

$$\varphi(x) = \varphi_0 + \varphi_1 x + \varphi_2 x^2 + \varphi_3 x^3 + \sum_{k=4}^{l}\varphi_k(x - r_{k-3})^3 I(x > r_{k-3}), \quad (3.2.15)$$

其中 $-\infty<r_1<r_2<\cdots<r_{l-3}<\infty$.当然,如果 $r_1,\cdots,r_{l-3}$是未定值,也应看作未知参数.此外,为了保证(3.2.14)式有平稳解,需要对参数 $\varphi_0,\varphi_1,\cdots,\varphi_l$ 附加某些限制,这将在第4章中讨论.形如(3.2.14)的模型称为样条自回归模型.模型(3.2.14)的自回归函数 $\varphi(x)$(见(3.2.15)式)具有二次可微性.当 $p=1$ 时,它可以很好地逼近模型(3.2.1)中的非参数自回归函数.因此,适当地选取(3.2.14)式中的 $r_1,\cdots,r_{l-3}$及 l 的值,可以用样条自回归模型(3.2.14)逼近一阶非参数自回归模型(3.2.1).但是当 $p\geqslant 2$ 时,推广上述方法有本质的困难,这主要是由于样条函数被推广到高维时遇到了困难.

(三) 指数自回归模型(Exponential Autoregression)

指数自回归模型被定义为

$$x_t = \varphi_{00} + \sum_{k=1}^{p}(\varphi_{0k} + \varphi_{1k}e^{-\gamma x_{t-1}^2})x_{t-k} + \varepsilon_t, \quad (3.2.16)$$

其中$\{\varepsilon_t\}$为白噪声序列,满足(3.2.2)式,$\varphi_{00},\varphi_{0k},\varphi_{1k}(k=1,\cdots,p)$和 $\gamma>0$ 为未知参数,正整数 p 称为模型(3.2.16)的阶数.模型(3.2.16)简记为 EAR(p).

指数自回归模型还有如下的推广形式:

$$x_t=\varphi_{00}+\sum_{k=1}^{p}\left\{\varphi_{0k}+\left(\sum_{j=1}^{l}\varphi_{jk}x_{t-1}^{j-1}\right)e^{-\gamma x_{t-1}^2}\right\}x_{t-k}+\varepsilon_t,\quad(3.2.17)$$

其中$\{\varepsilon_t\}$与(3.2.16)式中的相同,$\varphi_{00},\varphi_{0k},\varphi_{jk}(j=1,\cdots,l;k=1,\cdots,p)$和$\gamma>0$为未知参数.

在模型(3.2.16)中,指数项$e^{-\gamma x_{t-1}^2}$还可以用$e^{-\gamma x_{t-k}^2}$代替,常数γ也可以用γ_k代替,从而引出模型(3.2.16)的另一种更广的形式:

$$x_t=\varphi_{00}+\sum_{k=1}^{p}\left\{\varphi_{0k}+\left(\sum_{j=1}^{l}\varphi_{jk}x_{t-k}^{j-1}\right)e^{-\gamma_k x_{t-k}^2}\right\}\times x_{t-k}+\varepsilon_t.\quad(3.2.18)$$

形如(3.2.16)式的指数自回归模型,当γ为指定数值时,不难看出它具有拟线性自回归模型(3.2.12)的形式.以$p=\gamma=1$为例,此时模型(3.2.16)为

$$x_t=\varphi_{00}+\varphi_{01}x_{t-1}+\varphi_{11}x_{t-1}e^{-x_{t-1}^2}+\varepsilon_t.$$

相当于模型(3.2.12)中的$s=2$,且

$$f_1(x_{t-1})=x_{t-1},\ f_2(x_{t-1})=x_{t-1}e^{-x_{t-1}^2}.$$

模型(3.2.16)还可写成如下形式:

$$\begin{aligned}x_t=&\varphi_{00}+\varphi_{01}x_{t-1}+\cdots+\varphi_{0p}x_{t-p}\\&+\left(\sum_{k=1}^{p}\varphi_{1k}x_{t-k}\right)e^{-\gamma x_{t-1}^2}+\varepsilon_t.\end{aligned}\quad(3.2.19)$$

(3.2.19)式右端第一项恰好是线性自回归部分,第二项是非线性自回归部分.当$x_{t-1}^2\to\infty$时,(3.2.19)式的非线性自回归部分趋于0.从而可以看出模型(3.2.16)与线性自回归模型的关系.

指数自回归模型是以机械振动方程离散化而得到的(参见文献[84]),它有一定的实际背景.

(四) 有理分式自回归模型

有理分式自回归模型被定义为

$$x_t=\frac{\varphi(x_{t-1},\cdots,x_{t-p})}{\theta(x_{t-1},\cdots,x_{t-p})}+\varepsilon_t,\quad(3.2.20)$$

其中φ和θ分别为$x_{t-1},\cdots,x_{t-p}$的有穷阶多元多项式,其系数为未知参数.这里略去了其繁琐的具体形式;$\{\varepsilon_t\}$为白噪声序列,且满足(3.2.2)式.

在理论研究方面,为了保证(3.2.20)式有平稳解,需要对多项式φ和θ附加某些限制,所以它们不能是任意的多项式.在模型(3.2.20)的使用中,其未知参数的个数不能太多,当模型(3.2.20)的阶数$p=1$时,φ和θ

都是一元多项式，此时模型(3.2.20)可写为

$$x_t = \frac{\varphi_0 + \varphi_1 x_{t-1} + \cdots + \varphi_s x_{t-1}^s}{\theta_0 + \theta_1 x_{t-1} + \cdots + \theta_q x_{t-1}^q} + \varepsilon_t. \tag{3.2.21}$$

当然，即使是这种一元有理分式自回归模型，对多项式 φ 和 θ 的某些限制仍是必要的. 一种最简单的限制是分子的多项式不能比分母的多项式的次数高，且与分母的多项式互质. 例如

$$x_t = \frac{\varphi_1 x_{t-1}}{1 + \theta_1 x_{t-1}^2} + \varepsilon_t, \tag{3.2.22}$$

其中 $\theta_1 \geqslant 0$ 和 φ_1 是未知参数. 如果在(3.2.22)式中 θ_1 是已知的，则它就成为拟线性自回归模型. 在第 4 章中将证明模型(3.2.22)有平稳解.

在 $p=1$ 时，模型(3.2.21)可以很好地近似非线性自回归模型(3.2.1). 但是与样条自回归模型(3.2.14)相比较，模型(3.2.14)在统计分析中容易处理. 在实用中，除非有明确的实际背景，才使用(3.2.21)这样的模型.

当 $p>1$ 时，由于多元多项式比较复杂，为保证模型(3.2.20)有平稳解，对多项式 φ 和 θ 的限制条件也比较复杂，且较难处理. 下面讨论一个二元有理分式自回归模型的例子，即

$$x_t = \frac{\varphi_1 x_{t-1} x_{t-2}}{c_1 x_{t-1}^2 + c_2 x_{t-2}^2} + \varepsilon_t. \tag{3.2.23}$$

将此模型的分子分母同乘以一个常数 c 并不受影响，即

$$x_t = \frac{c\varphi_1 x_{t-1} x_{t-2}}{cc_1 x_{t-1}^2 + cc_2 x_{t-2}^2} + \varepsilon_t.$$

由此可见，当 $\varphi_1 \neq 0$ 时，(3.2.23)式中的参数 φ_1、c_1 和 c_2 中有一个多余参数. 令 $\theta_1 = c_1/\varphi_1$，$\theta_2 = c_2/\varphi_1$，则(3.2.23)式化为

$$x_t = \frac{x_{t-1} x_{t-2}}{\theta_1 x_{t-1}^2 + \theta_2 x_{t-2}^2} + \varepsilon_t. \tag{3.2.24}$$

如果(3.2.23)式中 $\varphi_1 = 0$，则由(3.2.23)式确定的随机序列 $\{x_t\}$ 退化为白噪声序列. 此例提醒我们，在考虑参数型非线性自回归模型时，应避免引入多余的参数.

对模型(3.2.24)，一般要求参数 $\theta_1 > 0$，$\theta_2 > 0$. 因为此时有

$$\left| \frac{x_{t-1} x_{t-2}}{\theta_1 x_{t-1}^2 + \theta_2 x_{t-2}^2} \right| \leqslant \frac{1}{2\sqrt{\theta_1 \theta_2}} < \infty.$$

由第 4 章的结果知，此时(3.2.24)式有平稳解.

三、部分参数型自回归模型

在参数型自回归模型与非参数型自回归模型之间还有一种部分参数型自回归模型，部分参数模型在统计分析中又称半参数模型. 它们与本书中的部分参数型自回归模型相似，但又不完全相同. 时间序列的部分参数模型没有明确的统一形式，但可以暂时借用(3.2.7)的形式：

$$x_t = \varphi(x_{t-1}, \cdots, x_{t-p}; \boldsymbol{\theta}) + \varepsilon_t, \tag{3.2.25}$$

其中非线性函数 φ 被参数 $\boldsymbol{\theta}$ 部分地确定，还有一部分属于非参数型的未知函数. 下面介绍几种不同的部分参数型自回归模型.

（一）广义线性自回归模型

在近代非线性分析中，有一种广义线性回归模型(参见文献[78])，将这一模型推广到时间序列分析中，就产生了如下的广义线性自回归模型：

$$x_t = \varphi(\alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p}) + \varepsilon_t, \tag{3.2.26}$$

其中 $\boldsymbol{\theta} = (\alpha_1, \cdots, \alpha_p)^\tau$ 为未知参数，φ 为未知函数，$\{\varepsilon_t\}$ 为白噪声序列，满足(3.2.2)式. 在统计分析时，既要估计参数 $\boldsymbol{\theta}$，又要估计一元函数 φ. 无论是为了准确地定义模型(3.2.26)，还是为了唯一地确定参数 $\boldsymbol{\theta}$ 和函数 φ 的估计值，都应当对参数 $\boldsymbol{\theta}$ 加以一定的限制. 一种常用的限制是要求 $\boldsymbol{\theta}$ 满足

$$\|\boldsymbol{\theta}\|^2 = \alpha_1^2 + \cdots + \alpha_p^2 = 1, \tag{3.2.27}$$

且 $\boldsymbol{\theta}$ 的第一个非零分量是正的. 在这样的限制下，(3.2.26)式中的参数 $\boldsymbol{\theta}$ 和函数 φ 是唯一确定的. 否则，可能导致不同的 $(\boldsymbol{\theta}, \varphi)$ 对应着同一个模型(3.2.26).

模型(3.2.26)的优点是，其非参数型的未知函数 φ 是一元函数. 如前面的讨论可知，一元函数 φ 虽然是非参数型的，但它可用样条函数很好地逼近，而样条自回归又便于拟合. 参数向量 $\boldsymbol{\theta}$ 还能反映模型一定的结构信息，称为*最有趣方向*或*主要信息方向*. 就像多元统计分析中的主成分方向那样，这种方法被称为*降维技术*.

（二）降维非参数自回归模型

广义线性模型的进一步推广是如下的降维非参数自回归模型：

$$x_t = \varphi(\boldsymbol{\theta}_1^\tau \boldsymbol{x}_{t-1}, \boldsymbol{\theta}_2^\tau \boldsymbol{x}_{t-1}, \cdots, \boldsymbol{\theta}_s^\tau \boldsymbol{x}_{t-1}) + \varepsilon_t, \tag{3.2.28}$$

其中 $\{\varepsilon_t\}$ 为白噪声序列，满足(3.2.2)式，$\boldsymbol{\theta}_i = (\theta_{i1}, \cdots, \theta_{ip})^\tau (i = 1, \cdots, p)$，且每个 $\boldsymbol{\theta}_i$ 都满足条件(3.2.27)式，$\boldsymbol{x}_t = (x_t, \cdots, x_{t-p+1})^\tau$，$\varphi$ 是 s 元非参数型函数，一般约定 $s < p$.

如果(3.2.28)式的 φ 具有(3.2.3)式的形式，则 x_t 满足如下关系式：

$$x_t = \varphi_1(\boldsymbol{\theta}_1^\tau \boldsymbol{x}_{t-1}) + \cdots + \varphi_s(\boldsymbol{\theta}_s^\tau \boldsymbol{x}_{t-1}) + \varepsilon_t. \tag{3.2.29}$$

在统计分析这一模型时，可将近代非线性回归分析中的投影寻踪方法(Projection Pursuit)(见文献[50])引入时间序列分析中.

(三) 可加性半参数自回归模型

如果自回归模型(3.2.1)具有如下形式：

$$x_t = \varphi_0 + \varphi_1 x_{t-1} + \cdots + \varphi_p x_{t-p} + h(x_{t-1}, \cdots, x_{t-p}) + \varepsilon_t, \tag{3.2.30}$$

其中$\{\varepsilon_t\}$为白噪声序列，满足(3.2.2)式，$\boldsymbol{\theta} = (\varphi_0, \varphi_1, \cdots, \varphi_p)^\tau$ 为未知参数，h 为 p 元未知函数，则称(3.2.30)式为**可加性半参数自回归模型**.

模型(3.2.30)是从近代回归分析中的半参数模型(参见文献[64]和[93])引申而来的. 模型(3.2.19)在形式上与模型(3.2.30)相似，但前者右端的第二项仍是参数型的. 模型(3.2.30)并没有降维作用，所以在应用时应主要关心(3.2.30)式中的线性部分的统计分析.

§ 3.3　随机条件方差模型

在 § 3.2 中，我们讨论了由预报导出的自回归模型(3.1.6)中$\{e_t\}$为白噪声序列的情况. 本节我们将讨论$\{e_t\}$为平稳鞅差序列的情况. 具体地说，就是讨论对(3.1.6)式的$\{e_t\}$的条件方差

$$E\{e_t^2 | x_{t-1}, x_{t-2}, \cdots\} = s^2(x_{t-1}, x_{t-2}, \cdots) \tag{3.3.1}$$

作了某些假定而产生的模型. 这样定义的模型称为**随机条件方差模型**.

首先回忆(3.1.6)式的由来，对于平稳序列，(3.1.3)式总是有意义的，或者说等式无条件成立. 从(3.1.3)式到(3.1.4)式，是人为地假定，或者说是一种分类，又可以说是一种近似. 无论如何，(3.1.4)式代表着一类重要的时间序列，它满足模型(3.1.6). 根据性质 1.4.1 的(2)知，(3.1.6)式中的$\{e_t\}$为平稳鞅差序列. 如在 § 1.4 所讨论的那样，在研究预报问题时，不应当强加给平稳鞅差序列$\{e_t\}$更多额外的限制. 但是，在本章定义的各种模型及以后研究模型的概率结构性质时，必须对模型中的随机噪声给出尽可能明确的假定，例如$\{e_t\}$为白噪声序列、正态白噪声序列等，才有可能给出相应模型的精细结构. 而在研究预报问题时，仅涉及到模型条件平均的结构特征，不一定涉及更细致的概率结构. 以前关于新息序列的不同假定的差异对线性序列的研究，影响并不太大，但在非线性时间序列分析中，必须加以区别. 因为对新息序列的不同假定可引出不同的非线

性模型来.

另外需要指出的一点是，对于自回归模型(3.1.6)的新息序列$\{e_t\}$，即使有如下等式

$$E\{x_t|x_{t-1},x_{t-2},\cdots\}=E\{x_t|x_{t-1},\cdots,x_{t-p}\},$$
$$E\{e_t|x_{t-1},x_{t-2},\cdots\}=E\{e_t|x_{t-1},\cdots,x_{t-p}\}=0$$

成立，仍然不可以断言

$$E\{e_t^2|x_{t-1},x_{t-2},\cdots\}=E\{e_t^2|x_{t-1},\cdots,x_{t-p}\}$$

成立. 例如，考虑如下的非线性模型：

$$x_t=\alpha\varepsilon_t|\varepsilon_{t-1}|^{1/2}, \tag{3.3.2}$$

式中$\{\varepsilon_t\}$为正态白噪声序列，且$\varepsilon_t\sim N(0,1)$. 易见

$$\begin{aligned}E\{x_t|x_{t-1},x_{t-2},\cdots\}&=\alpha E\{\varepsilon_t|\varepsilon_{t-1}|^{1/2}|x_{t-1}\}\\&=\alpha(E\varepsilon_t)E\{|\varepsilon_{t-1}|^{1/2}|x_{t-1}\}\\&=0.\end{aligned}$$

可见(3.3.2)式的x_t为平稳鞅差序列，也为特殊的LAR(1)序列，即其自回归系数为0值，也即

$$x_t=0\cdot x_{t-1}+e_t,$$

其中$e_t=\alpha\varepsilon_t|\varepsilon_{t-1}|^{1/2}$. 下面计算$e_t$的条件方差. 先计算$E\{e_t^2|x_{t-1}\}$. 注意到$\{\varepsilon_t\}$是独立同分布的，故

$$\begin{aligned}E\{e_t^2|x_{t-1}\}&=E\{\alpha^2\varepsilon_t^2|\varepsilon_{t-1}||x_{t-1}\}\\&=\alpha^2E\{|\varepsilon_{t-1}||x_{t-1}\}\\&=\alpha^2E\{|\varepsilon_{t-1}||\varepsilon_{t-1}|\varepsilon_{t-2}|^{1/2}\}.\end{aligned} \tag{3.3.3}$$

再计算$E\{e_t^2|x_{t-1},x_{t-2},\cdots\}$，容易验证

$$\begin{aligned}\sigma(x_s,s\leqslant t-1)&=\sigma(\alpha\varepsilon_s|\varepsilon_{s-1}|^{1/2},s\leqslant t-1)\\&\supset\sigma(|\varepsilon_s||\varepsilon_{s-1}|^{1/2},\ s\leqslant t-1)\\&=\sigma\left(\log|\varepsilon_s|+\frac{1}{2}\log|\varepsilon_{s-1}|,s\leqslant t-1\right).\end{aligned}$$

记$\eta_s=\log|\varepsilon_s|$，显然$\{\eta_s\}$为独立同分布序列，且

$$y_s=\eta_s+\frac{1}{2}\eta_{s-1}$$

为可逆的MA(1)序列，从而又有

$$\begin{aligned}\sigma(x_s,s\leqslant t-1)\supset\sigma\left(\eta_s+\frac{1}{2}\eta_{s-1},\ s\leqslant t-1\right)\supset\sigma(\eta_s,s\leqslant t-1)\\=\sigma(|\varepsilon_s|,s\leqslant t-1).\end{aligned}$$

注意到$|\varepsilon_{t-1}|$是$\sigma\left(\frac{1}{2}\eta_{s-1}+\eta_s, s\leqslant t-1\right)$可测的，从而也是$\sigma(x_s, s\leqslant t-1)$可测的，于是有

$$E\{e_t^2 | x_{t-1}, x_{t-2}, \cdots\} = \alpha^2 E\{|\varepsilon_{t-1}| \,|\, x_{t-1}, x_{t-2}, \cdots\}$$
$$= \alpha^2 |\varepsilon_{t-1}|.$$

在正态条件下，用直接验证可知

$$E\{|\varepsilon_{t-1}| \,\big|\, |\varepsilon_{t-1}|\varepsilon_{t-2}|^{1/2}\} \not\equiv |\varepsilon_{t-1}|,$$

从而

$$E\{e_t^2 \,\big|\, x_{t-1}\} \not\equiv E\{e_t^2 \,\big|\, x_{t-1}, x_{t-2}, \cdots\}.$$

一、自回归条件异方差模型

如果时间序列$\{x_t\}$服从如下自回归模型

$$x_t = \varphi(x_{t-1}, \cdots, x_{t-p}) + e_t, \tag{3.3.4}$$

且新息序列$\{e_t\}$满足

$$e_t | \mathscr{F}_{t-1} \sim N(0, h_t), \tag{3.3.5}$$

其中$\mathscr{F}_{t-1}=\sigma(x_{t-1}, x_{t-2}, \cdots)$为由$\{x_{t-1}, x_{t-2}, \cdots\}$生成的$\sigma$-域，(3.3.5)式表示在给定$\mathscr{F}_{t-1}$时，$e_t$的条件分布为正态$N(0, h_t)$. 如果条件方差$h_t$满足如下模型

$$h_t = h(e_{t-1}, \cdots, e_{t-q}), \tag{3.3.6}$$

其中h是一个q元非负可测函数，则称$\{x_t\}$服从自回归条件异方差(Autoregression Conditional Heteroskedasticity)模型，并简记为ARCH(p, q)，其中p为模型(3.3.4)的阶数，q为模型(3.3.6)的阶数. 如果h_t满足如下模型

$$h_t = h(e_{t-1}, \cdots, e_{t-q}, h_{t-1}, \cdots, h_{t-r}), \tag{3.3.7}$$

其中h为$q+r$元非负可测函数，则称$\{x_t\}$服从广义自回归条件异方差(Generalized ARCH)模型，并简记为GARCH(p, q, r).

无论是对ARCH模型，还是对GARCH模型，原则上§3.2中的非参数、参数和半参数分类法对它们都适用. 在这里就不作这样单调的重复叙述了，只是对条件方差h_t引入一种有用的线性参数模型，即

$$h_t = \alpha_0 + \alpha_1 e_{t-1}^2 + \cdots + \alpha_q e_{t-q}^2 + \theta_1 h_{t-1} + \cdots + \theta_r h_{t-r}, \tag{3.3.8}$$

式中系数满足$\alpha_i \geqslant 0 \quad (i=0,1,\cdots,q)$，$\theta_j \geqslant 0 \quad (j=1,\cdots,r)$.

另外，如果e_t具有条件正态分布，那么(3.3.5)式也可被如下的推广

方式所代替,即

$$e_t = \varepsilon_t h_t^{1/2}, \tag{3.3.9}$$

式中$\{\varepsilon_t\}$为正态白噪声序列,且满足

$$\begin{cases} E\varepsilon_t = 0,\ E\varepsilon_t^2 = 1; \\ \varepsilon_t \text{ 与 } \{h_s, s < t\} \text{ 相互独立}. \end{cases} \tag{3.3.10}$$

注意,(3.3.10)式中不用假定$E\varepsilon_t^2=\sigma^2$,而用$E\varepsilon_t^2=1$,其原因在于:如果假定$E\varepsilon_t^2=\sigma^2$,则由(3.3.9)式,有

$$Ee_t^2 = E\varepsilon_t^2 Eh_t = \sigma^2 Eh_t.$$

这样,σ^2和h_t在变换σ^2/c和ch_t $(c>0)$下不影响模型. 为了模型的唯一确定,我们限定$E\varepsilon_t^2=1$.

将(3.3.4)、(3.3.8)和(3.3.9)三式联合起来,便得到如下关于$\{x_t\}$的GARCH模型:

$$\begin{cases} x_t=\varphi(x_{t-1},\cdots,x_{t-p})+e_t, \\ e_t=\varepsilon_t h_t^{1/2}, \\ h_t=\alpha_0+\alpha_1 e_{t-1}^2+\cdots+\alpha_q e_{t-q}^2+\theta_1 h_{t-1}+\cdots+\theta_r h_{t-r}. \end{cases} \tag{3.3.11}$$

模型(3.3.11)的一个有用的特例是φ为线性函数的情况,即

$$\begin{cases} x_t=\varphi_0+\varphi_1 x_{t-1}+\cdots+\varphi_p x_{t-p}+e_t, \\ e_t=\varepsilon_t h_t^{1/2}, \\ h_t=\alpha_0+\alpha_1 e_{t-1}^2+\cdots+\alpha_q e_{t-q}^2+\theta_1 h_{t-1}+\cdots+\theta_r h_{t-r}, \end{cases} \tag{3.3.12}$$

式中自回归系数$\varphi_1,\cdots,\varphi_p$满足条件(1.2.5)式. 简记(3.3.12)式为LGARCH(p,q,r)模型. 模型(3.3.12)的一个最简单的特例是

$$\begin{cases} x_t = \varphi_0 + \varphi_1 x_{t-1} + \cdots + \varphi_p x_{t-p} + e_t, \\ e_t = \varepsilon_t h_t^{1/2}, \\ h_t = \alpha_0 + \alpha_1 e_{t-1}^2. \end{cases} \tag{3.3.13}$$

模型(3.3.12)有着广泛的实际背景. 以模型(3.3.13)为例,从预报的角度来看,模型(3.3.13)的一步预报误差e_t的条件方差正比于前一时刻的误差e_{t-1}^2,而不是常数. 更一般地,模型(3.3.12)在t时刻的一步预报误差e_t的条件方差是t时刻以前的误差$e_{t-s}^2(s>0)$的函数,而不是常数. 在经济领域中,许多经济现象——诸如通货膨胀率、兑换率、股票价值等,都具有这样的性质. 所以ARCH模型和GARCH模型在经济领域中倍受重视(参见文献[14]和[45]). 此外,在许多其他领域中也有类似的应用背景. 例如在气象、水文等领域中.

二、β-ARCH 模型和 β-GARCH 模型

对 ARCH 模型或 GARCH 模型，为保证其平稳性，对参数的限制极苛刻. 例如对 LARCH$(p,1)$模型(3.3.13)，将其第二个方程两端平方并递推，可得

$$
\begin{aligned}
e_t^2 &= \varepsilon_t^2(\alpha_0 + \alpha_1 e_{t-1}^2) = \alpha_0\varepsilon_t^2 + \alpha_1\varepsilon_t^2 e_{t-1}^2 \\
&= \alpha_0\varepsilon_t^2 + \alpha_1\varepsilon_t^2(\alpha_0\varepsilon_{t-1}^2 + \alpha_1\varepsilon_{t-1}^2 e_{t-2}^2) \\
&= \alpha_0\varepsilon_t^2 + \alpha_0\alpha_1\varepsilon_t^2\varepsilon_{t-1}^2 + \alpha_1^2\varepsilon_t^2\varepsilon_{t-1}^2 e_{t-2}^2 \\
&= \cdots\cdots \\
&= \alpha_0\sum_{j=0}^{\infty}\alpha_1^j\prod_{k=0}^{j}\varepsilon_{t-k}^2.
\end{aligned}
$$

容易验证$\{e_t^2\}$是平稳序列而且 $Ee_t^2<\infty$，当且仅当 $0<\alpha_1<1$. 为了对模型(3.3.13)进行统计分析，往往还要求它具有高于 2 阶的矩存在，这时对 α_1 的限制会更加苛刻. 对 LARCH(p,q)模型，即模型(3.3.12)中 $\theta_j=0(j=1,\cdots,r)$的情形，由第 4 章的讨论知，为使该模型有平稳解，系数 $\alpha_1,\cdots,\alpha_q$ 必须满足

$$
\alpha_i \geqslant 0 \quad (i=1,\cdots,q), \quad \sum_{j=1}^{q}\alpha_j < 1.
$$

对系数 α_i 的这些限制是偏强的，为了放宽对上述系数的限制，我们引入如下的随机条件方差模型：如果时间序列$\{x_t\}$服从如下随机条件方差模型

$$
\begin{cases}
x_t = \varphi(x_{t-1},\cdots,x_{t-p}) + e_t, \\
e_t = \varepsilon_t h_t^{1/2}, \\
h_t = h(e_{t-1},\cdots,e_{t-q},h_{t-1},\cdots,h_{t-r}),
\end{cases}
\tag{3.3.14}
$$

式中$\{\varepsilon_t\}$为白噪声序列，满足(3.3.10)式，且可测函数 h 具有如下形式：

$$
h_t = \alpha_0 + \alpha_1|e_{t-1}|^{2\beta} + \cdots + \alpha_q|e_{t-q}|^{2\beta}, \tag{3.3.15}
$$

其中 $\alpha_i\geqslant 0\ \ (i=0,1,\cdots,q)$，$0\leqslant\beta<1$，则称$\{x_t\}$服从 β-ARCH$(p,q)$模型. 如果可测函数 h 具有如下形式：

$$
\begin{aligned}
h_t = \alpha_0 + \alpha_1|e_{t-1}|^{2\beta_1} + \cdots + \alpha_q|e_{t-q}|^{2\beta_1} \\
+ \theta_1 h_{t-1}^{\beta_2} + \cdots + \theta_r h_{t-r}^{\beta_2},
\end{aligned}
\tag{3.3.16}
$$

其中 $\alpha_i\geqslant 0\ \ (i=0,1,\cdots,q)$，$\theta_j\geqslant 0\ \ (j=1,\cdots,r)$，$0\leqslant\beta_1<1,0\leqslant\beta_2\leqslant 1$，则称$\{x_t\}$服从 β-GARCH$(p,q,r)$模型.

在第 4 章中我们将会看到，在(3.3.15)式中，只要 $0\leqslant\beta<1$，对其系数 $\alpha_1,\cdots,\alpha_q$ 除了非负条件外，不加任何其他的限制，β-ARCH(p,q)模型

(3.3.15)式必有平稳解.

三、函数型随机条件方差自回归模型

在前两种随机条件方差模型中，预报误差 e_t 的条件方差 h_t 都是 e_{t-1}^2, $e_{t-2}^2,\cdots$ 的函数，或者是 $e_{t-1}^2,e_{t-2}^2,\cdots$ 和 $h_{t-1},h_{t-2},\cdots$ 的函数. 下面介绍另一种随机条件方差模型，它是将预报误差 e_t 的条件方差直接表达成原序列 $x_{t-1},x_{t-2},\cdots,x_{t-p}$ 的函数，即

$$\begin{aligned}E\{e_t^2\,|\,x_{t-1},x_{t-2},\cdots\} &= E\{e_t^2\,|\,x_{t-1},\cdots,x_{t-p}\}\\ &= S^2(x_{t-1},\cdots,x_{t-p}),\end{aligned}\tag{3.3.17}$$

或者说，假定 e_t 的条件方差只与 $x_{t-1},\cdots,x_{t-p}$ 有关. 为了理论研究和统计分析的方便，不仅假定(3.3.17)式成立，还进一步假定

$$e_t = \varepsilon_t S(x_{t-1},\cdots,x_{t-p}),\tag{3.3.18}$$

式中 $\{\varepsilon_t\}$ 为白噪声序列，满足

$$\begin{cases}E\varepsilon_t = 0, \quad E\varepsilon_t^2 = 1;\\ \varepsilon_t \text{ 与 } \{x_s, s < t\} \text{ 相互独立},\end{cases}\tag{3.3.19}$$

$S(x_1,\cdots,x_p)$ 是 p 元非负可测函数.

显然，由(3.3.18)式定义的 e_t 必满足(3.3.17)式.

如果时间序列 $\{x_t\}$ 满足如下随机条件方差模型

$$x_t = \varphi(x_{t-1},\cdots,x_{t-p}) + \varepsilon_t S(x_{t-1},\cdots,x_{t-p}),\tag{3.3.20}$$

其中 φ 和 S 都是 p 元可测函数，且 S 是非负函数，$\{\varepsilon_t\}$ 为白噪声序列，满足条件(3.3.19)式，则称 $\{x_t\}$ 服从 **p 阶函数型随机条件方差自回归模型**.

在模型(3.3.20)中，p 元可测函数 φ 和 S 可以按§3.2的非参数、参数型等各种假定分类. 这里不再一一赘述了.

在本节开始时，我们曾以反例说明：即使是 p 阶自回归模型的新息序列 $\{e_t\}$，其条件方差 $S^2(x_{t-1},x_{t-2},\cdots)$ 也不一定仅为 $x_{t-1},\cdots,x_{t-p}$ 的函数. 所以说，(3.3.17)式是一种人为的假定，或者说是一种近似. 总之，(3.3.17)式可作为新息序列 $\{e_t\}$ 的一种特殊类型来研究. 现在考虑(3.3.17)式的 e_t 是否总能表示成(3.3.18)式的问题. 这里只讨论 $S(x_{t-1},\cdots,x_{t-p})$ 为正值函数的情况. 令

$$\eta_t = e_t/S(x_{t-1},\cdots,x_{t-p}),$$

则有

$$E\{\eta_t\,|\,x_{t-1},x_{t-2},\cdots\}$$

$$= \{S(x_{t-1},\cdots,x_{t-p})\}^{-1}E\{e_t | x_{t-1},x_{t-2},\cdots\} = 0, \text{ a.s.},$$

$$E\{\eta_t^2 | x_{t-1},x_{t-2},\cdots\}$$

$$= \{S(x_{t-1},\cdots,x_{t-p})\}^{-2}E\{e_t^2 | x_{t-1},x_{t-2},\cdots\}$$

$$= 1, \text{ a.s.}.$$

以上两式表明$\{\eta_t\}$是一平稳鞅差序列，且具有常数条件方差. 于是有

$$e_t = \eta_t S(x_{t-1},\cdots,x_{t-p}). \tag{3.3.21}$$

此式表明，在(3.3.17)式的前提下，e_t 总可以表示为(3.3.21)式的形式. (3.3.21)式在形式上与(3.3.18)式相同，但其因子$\{\eta_t\}$不是白噪声序列. 所以说，尽管(3.3.18)式仍是人为的假定，但须在这一假定下，关于 e_t 的模型界定才够精确而便于研究.

由以上讨论可知，函数型随机条件方差模型反映了新息 e_t 的条件方差与临近时刻的时间序列的取值有密切的关系. 事实上，在门限自回归模型(3.2.9)中，当 $x_{t-d}\in R_j$ 时，如果将 ε_t 换成 ε_{it}，这时的门限自回归模型的新息也反映了上述现象. 具体地说，将(3.2.9)式改写为

$$x_t = \sum_{j=1}^{l}\left\{\varphi_{j0} + \sum_{k=1}^{p}\varphi_{jk}x_{t-k}\right\}I(r_j \leqslant x_{t-d} < r_{j+1})$$

$$+ \sum_{j=1}^{l}\varepsilon_{jt}I(r_j \leqslant x_{t-d} < r_{j+1}), \tag{3.3.22}$$

其中$\{\varepsilon_{jt}\}$($1\leqslant j\leqslant l$)是相互独立的白噪声序列，满足

$$E\varepsilon_{jt} = 0, \quad E\varepsilon_{jt}^2 = \sigma_j^2 < \infty \quad (j = 1,\cdots,l). \tag{3.3.23}$$

且 ε_{jt}($1\leqslant j\leqslant l$)与$\{x_s, s<t\}$相互独立. 由(3.3.22)式可知：

$$E\{x_t | x_{t-1},\cdots,x_{t-p}\}$$

$$= \sum_{j=1}^{l}\left\{\varphi_{j0} + \sum_{k=1}^{p}\varphi_{jk}x_{t-k}\right\}I(r_j \leqslant x_{t-d} < r_{j+1}),$$

$$e_t = x_t - E\{x_t | x_{t-1},\cdots,x_{t-p}\}$$

$$= \sum_{j=1}^{l}\varepsilon_{jt}I(r_j \leqslant x_{t-d} < r_{j+1}),$$

$$E\{e_t^2 | x_{t-1},\cdots,x_{t-p}\}$$

$$= E\left\{\left[\sum_{j=1}^{l}\varepsilon_{jt}I(r_j \leqslant x_{t-d} < r_{j+1})\right]^2 \middle| x_{t-1},\cdots,x_{t-p}\right\}$$

$$= \sum_{j=1}^{l}\sigma_j^2 I(r_j \leqslant x_{t-d} < r_{j+1}).$$

上式表明，对(3.3.22)式确定的时间序列，其预报误差的条件方差与时间

序列的临近时刻的取值有关. 易见,模型(3.3.22)中的误差项 ε_{jt}的方差可随 x_{t-d}的值域不同而不同. 因此,模型(3.3.22)可以较好地描述实际情况,这也正是门限自回归模型能被广泛应用于实际的原因之一. 例如,在气象领域中,许多气象要素的记录序列当其取值水平较高时,相应的新息的条件方差也较大.

§ 3.4　非线性自回归滑动平均模型

前两节中介绍的时间序列模型,都是基于(3.1.6)式的自回归预报模型,即时间序列的一步预报只与有限个历史值有关. 对于一步预报与无穷多个历史值有关的情况,我们很难基于(3.1.3)式的无穷多元函数 $\varphi(x_{t-1},x_{t-2},\cdots)$进行分类和研究. 因此,本节和 § 3.5 所讨论的模型是直接提出各种参数模型,它们有的是从线性自回归滑动平均模型的推广而来的,有的是从其他实际背景提出来的. 一般说来,这些模型的一步预报不满足(3.1.4)式. 本节介绍的非线性自回归滑动平均模型的一般形式为

$$x_t = \varphi(x_{t-1},\cdots,x_{t-p},\varepsilon_t,\varepsilon_{t-1},\cdots,\varepsilon_{t-q}), \qquad (3.4.1)$$

其中$\{\varepsilon_t\}$为白噪声序列,满足(3.2.2)式. 形如(3.4.1)的 NLARMA(p,q)模型还有以下特殊情况:

$$x_t = \varphi(x_{t-1},\cdots,x_{t-p},\varepsilon_{t-1},\cdots,\varepsilon_{t-q}) + \varepsilon_t, \qquad (3.4.2)$$

或者更特殊的情况:

$$x_t = \varphi(x_{t-1},\cdots,x_{t-p}) + \theta(\varepsilon_{t-1},\cdots,\varepsilon_{t-q}) + \varepsilon_t. \qquad (3.4.3)$$

易见,熟知的 LARMA(p,q)模型(1.2.4)是模型(3.4.1)的特例,同样也是模型(3.4.2)和(3.4.3)的特例. 于是从线性到非线性推广时,又出现了以上三种形式. 容易想象,即使是最简单的情况(3.4.3)式,也比 LARMA 模型复杂得难以估计. 因此,一般都很少讨论(3.4.1)～(3.4.3)形式的 NLARMA 模型,而是研究其某些特殊的参数型 NLARMA 模型或半参数型的 NLARMA 模型. 另外,当 $q=0$ 时,(3.4.3)式即为(3.2.1)式的 NLAR 模型,故以下只讨论 $q>0$ 的情况.

一、双线性模型

如果时间序列$\{x_t\}$满足如下关系式:

$$x_t = \sum_{j=1}^{p}\varphi_j x_{t-j} + \sum_{k=0}^{q}\theta_k\varepsilon_{t-k} + \sum_{i=1}^{Q}\sum_{l=1}^{P}\beta_{il}x_{t-l}\varepsilon_{t-i}, \qquad (3.4.4)$$

其中$\{\varepsilon_t\}$为白噪声序列，满足(3.2.2)式，则称$\{x_t\}$服从双线性模型(Bilinear Model)，简记为 BL(p,q,P,Q)，其中(p,q,P,Q)为模型(3.4.4)的阶数.

对双线性模型(3.4.4)研究得较多的是以下两种特殊情况：一种情形是

$$x_t = \sum_{k=1}^{p}\varphi_k x_{t-k} + \sum_{i=1}^{Q}\sum_{l=1}^{P}\beta_{il}x_{t-l}\varepsilon_{t-i} + \varepsilon_t, \qquad (3.4.5)$$

另一种情形是

$$x_t = \sum_{j=1}^{q}\theta_j\varepsilon_{t-j} + \sum_{i=1}^{Q}\sum_{l=1}^{P}\beta_{il}\varepsilon_{t-l}\varepsilon_{t-i} + \varepsilon_t. \qquad (3.4.6)$$

模型(3.4.6)是线性滑动平均模型的二阶推广，也是第 1 章中提到的 Volterra 展式的二阶近似. 模型(3.4.6)的优点是，在$\{\varepsilon_t\}$为白噪声的假定下$\{x_t\}$总是平稳的. 但是关于此模型的统计建模和预报方法的研究比较困难. 关于这些模型的研究可参见文献[54]和[51].

二、逆自回归模型

在近代高维数据的回归分析研究中，有一种被称为逆自回归的方法(见文献[73])，它是高维数据分析中的一种降维方法. 把这一方法引入时间序列分析中，就给出了如下的一种新的非线性时间序列模型：

$$x_t = \varphi(\alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + \varepsilon_t), \qquad (3.4.7)$$

式中φ为一元未知函数，$\boldsymbol{\theta}=(\alpha_1,\cdots,\alpha_p)^\tau$为$p$维未知参数向量，它满足广义线性自回归模型(3.2.26)中的约定(3.2.27)式，$\{\varepsilon_t\}$为白噪声序列，满足(3.2.2)式.

与(3.4.7)式联系密切的还有另一种模型

$$f(x_t) = \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + \varepsilon_t, \qquad (3.4.8)$$

式中f为一元未知函数，参数向量$\boldsymbol{\theta}=(\alpha_1,\cdots,\alpha_p)^\tau$和$\{\varepsilon_t\}$的假定如模型(3.4.7)所述. 显然(3.4.8)式已超出本节所定义的 NLARMA 模型. 实际上，它是如下更广义的时间序列模型的特例，即

$$\Psi(x_t, x_{t-1}, \cdots, x_{t-p}) = \varepsilon_t. \qquad (3.4.9)$$

对模型(3.4.9)，在此不作进一步的讨论.

在逆自回归分析中，最使人感兴趣的情况是函数f为可逆的，即存在函数$\varphi = f^{-1}$为f的逆函数. 于是模型(3.4.7)和模型(3.4.8)是等价的模型. 在逆回归分析中也有类似的情况. 在此情况下，研究模型(3.4.8)的

统计方法变得比较容易了.当然,在时间序列分析中,即使 f 和 φ 不可逆,模型(3.4.7)和模型(3.4.8)都有独立研究的价值.因为它们属于半参数模型,而且未知函数是一维的.如前所述,一维函数可用拟线性自回归模型来逼近,又容易在计算机上显示图形,有助于统计分析.但是,当 f 不可逆时,研究模型(3.4.8)的平稳性尚未见之于文献.

§3.5 其他非线性时间序列模型

在前面几节中介绍了多种不同类型的非线性时间序列模型.用前几节的模型进行新的组合,就可以产生许多新的模型.总之,非线性模型不像线性模型那样有统一的表达方式.没有一种包括所有非线性模型而又便于理论和统计处理的统一的形式存在,即使对有限参数的非线性模型也不能做到这一点.因此,对非线性时间序列的研究只能对不同类型的模型分别进行研究.某些类型的模型可能在某些领域中有较好的应用背景,而在另外的领域中却不适合.在本节中再来讨论两类时间序列模型,它们不能被前面讨论的模型所包括.

一、双重线性模型

为了克服线性自回归模型的局限性,拓广它的描述实际数据序列的能力,又尽可能避免引入非线性模型,有一种方法是引入随机系数的自回归模型,即

$$x_t = \varphi_0 + \sum_{k=1}^{p}[\varphi_k + \beta_k(t)]x_{t-k} + \varepsilon_t, \qquad (3.5.1)$$

其中 $\varphi_0,\varphi_1,\cdots,\varphi_p$ 为未知参数, $\{\varepsilon_t\}$ 为白噪声序列, 满足(3.2.2)式, $\beta_1(t),\cdots,\beta_p(t)$ 为 p 个随机序列.显然,当 $\beta_k(t)=0$, a.s. $(k=1,\cdots,p)$ 时,(3.5.1)式就成为通常的LAR(p)模型(1.3.7)式.当 $\beta_k(t)=\varepsilon_{t-k}(k=1,\cdots,p)$ 时,(3.5.1)式成为双线性模型(3.4.4)的特殊情况.

在一般情况下, $\beta_k(t)$ $(k=1,\cdots,p)$ 可以假定为平稳序列,在适当条件下, $\{x_t\}$ 也具有平稳解.但是对此类模型的统计建模比较困难,因此在文献中又研究了 $\beta_k(t)$ $(k=1,\cdots,p)$ 满足适当的LARMA模型的情况,即研究如下的随机模型:

$$
\begin{cases}
x_t = \varphi_0 + \sum_{k=1}^{p} [\varphi_k + \beta_k(t)] x_{t-k} + \varepsilon_t, \\
\beta_k(t) = \alpha_{k1}\beta_k(t-1) + \cdots + \alpha_{ks_k}\beta_k(t-s_k) \\
\qquad + \theta_{k1}\eta_k(t-1) + \cdots + \theta_{kr_k}\eta_k(t-r_k) \\
\qquad + \eta_k(t)(k=1,\cdots,p),
\end{cases}
\tag{3.5.2}
$$

式中$\{\eta_k(t)\}$ $(k=1,\cdots,p)$为 p 个相互独立的白噪声序列，且与$\{\varepsilon_t\}$独立，φ_0、φ_k、α_{ki}、θ_{kj} $(k=1,\cdots,p;i=1,\cdots,s_k;j=1,\cdots,r_k)$为未知参数. 换言之，每个$\{\beta_k(t)\}$都是一个 LARMA$(s_k,r_k)$序列且相互独立. (3. 5. 2)式称为**双重线性自回归模型**.

模型(3. 5. 2)太复杂了，不利于实际应用，而且研究其结构性质也很困难. 实际上，在文献中讨论较多的是它的一些特殊情况. 其中一种特殊情况是

$$
\begin{cases}
x_t = \varphi_0 + [\varphi_1 + \beta(t)] x_{t-1} + \varepsilon_t, \\
\beta(t) = \alpha\beta(t-1) + \eta_t.
\end{cases}
\tag{3.5.3}
$$

模型(3. 5. 3)称为 AR(1)-AR(1)**双重线性模型**.

另一种特殊情况是

$$
\begin{cases}
x_t = \varphi_0 + [\varphi_1 + \beta(t)] x_{t-1} + \varepsilon_t, \\
\beta(t) = \eta_t + \theta_1\eta_{t-1} + \cdots + \theta_r\eta_{t-r}.
\end{cases}
\tag{3.5.4}
$$

模型(3. 5. 4)称为 AR(1)-MA(r)**双重线性模型**.

二、非线性多元时间序列模型

本书不多涉及多元时间序列模型，这里仅作简单的介绍.

一阶 m 维非线性自回归模型定义为

$$
\boldsymbol{x}_t = \boldsymbol{\varphi}(\boldsymbol{x}_{t-1}) + \boldsymbol{\varepsilon}_t,
\tag{3.5.5}
$$

其中 $\boldsymbol{x}_t = (x_{1t},\cdots,x_{mt})^\tau$, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t},\cdots,\varepsilon_{mt})^\tau$,

$\boldsymbol{\varphi}$ 是从 R^m 到 R^m 的可测函数，$\{\boldsymbol{\varepsilon}_t\}$为 m 维向量白噪声序列，满足

$$
\begin{cases}
E\boldsymbol{\varepsilon}_t = 0, \ E\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\tau = \boldsymbol{\Gamma}, \\
\boldsymbol{\varepsilon}_t \text{ 与 } \{\boldsymbol{x}_s, s < t\} \text{ 相互独立}.
\end{cases}
\tag{3.5.6}
$$

类似于第 2 章中化 p 阶一维动态系统为一阶 p 维动态系统的方法，可将(3. 3. 20)式定义的 NLAR(p)模型化为(3. 5. 5)式的形式. 令

$$
\boldsymbol{\varepsilon}_t = \varepsilon_t \boldsymbol{u}, \quad \text{其中 } \boldsymbol{u} = (1,0,\cdots,0)^\tau,
$$

$$
\boldsymbol{\varphi}(\boldsymbol{x}_{t-1}) = (\varphi(\boldsymbol{x}_{t-1}), x_{t-1},\cdots,x_{t-p+1})^\tau,
$$

则(3.2.20)式可写成

$$x_t = \varphi(x_{t-1}) + S(x_{t-1})\varepsilon_t. \tag{3.5.7}$$

特别当(3.3.20)式中 $S(x_{t-1})\equiv 1$ 时,(3.3.20)式成为(3.2.1)式,而相应的(3.5.7)式成为(3.5.5)式的形式,即

$$x_t = \varphi(x_{t-1}) + \varepsilon_t u. \tag{3.5.8}$$

虽然形如(3.2.1)式的 NLAR(p)模型可以改写成(3.5.8)式的形式,但(3.5.8)式只是(3.5.5)式的一种特殊情况,在(3.5.8)式中$\boldsymbol{\varepsilon}_t=\varepsilon_t \boldsymbol{u}$,且

$$E\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^{\tau} = E\varepsilon_t^2 \boldsymbol{u}\boldsymbol{u}^{\tau} = \sigma^2 \boldsymbol{u}\boldsymbol{u}^{\tau}$$
$$= \sigma^2 \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}. \tag{3.5.9}$$

协方差矩阵(3.5.9)的秩是 1($\sigma^2>0$ 时),而在(3.5.5)式中的 $\boldsymbol{\varepsilon}_t$ 的协方差矩阵 Γ 可为满秩正定阵. 本书所涉及的多元时间序列模型主要是形如(3.5.8)式的非线性模型,而很少讨论形如(3.5.5)式的非线性模型,由于篇幅所限,不讨论一般多元时间序列. 当然,这并不是说非线性多元时间序列与一元的情况在概率结构和统计分析上没有本质差异,实际情况恰恰相反. 这里不介绍多元的情况,除了篇幅所限以外,还由于对它的研究所涉及的数学知识更广更深. 这里举两个最简单的二元非线性时间序列模型的例子,以示其特性.

例 3.5.1 考虑如下一阶二维时间序列模型:

$$\begin{cases} \boldsymbol{x}_t = \begin{pmatrix} 0.5 & -0.5 \\ 0.5 & 0.5 \end{pmatrix} \boldsymbol{x}_{t-1} + \varepsilon_t \begin{pmatrix} 1 \\ 1 \end{pmatrix}; \\ \boldsymbol{x}_0 = (1,1)^{\tau}. \end{cases} \tag{3.5.10}$$

式中$\{\varepsilon_t\}$为白噪声序列,$P(\varepsilon_t=1)=P(\varepsilon_t=-1)=\frac{1}{2}$. 模型(3.5.10)的样本轨道为二维序列$\{\boldsymbol{x}_t\}$,其样本轨道的图形如一片雪花,如图 3.5.1 所示.

例 3.5.2 考虑如下的一阶二维时间序列模型:

$$\begin{cases} \boldsymbol{x}_t = \sum_{j=1}^{4} \{A_j \boldsymbol{x}_{t-1} + \boldsymbol{b}_j\} I(\varepsilon_t = j); \\ \boldsymbol{x}_0 = (1,1)^{\tau}. \end{cases} \tag{3.5.11}$$

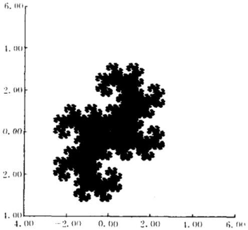


图 3.5.1　模型(3.5.10)的样本轨道



图 3.5.2　模型(3.5.11)的样本轨道

其中$\{\varepsilon_t\}$为白噪声序列，$P(\varepsilon_t=1)=0.01$，$P(\varepsilon_t=2)=0.85$，$P(\varepsilon_t=3)=P(\varepsilon_t=4)=0.07$,系数矩阵和系数向量分别为

$$A_1=\begin{pmatrix}0 & 0\\ 0 & 0.25\end{pmatrix},\qquad A_2=\begin{pmatrix}0.85 & 0.04\\ -0.04 & 0.85\end{pmatrix},$$

$$A_3=\begin{pmatrix}0.20 & -0.20\\ 0.26 & 0.22\end{pmatrix},\qquad A_4=\begin{pmatrix}-0.15 & 0.28\\ 0.28 & 0.24\end{pmatrix},$$

$$\boldsymbol{b}_1=(0,0)^{\tau},\qquad \boldsymbol{b}_2=(0,1.6)^{\tau},$$

$$\boldsymbol{b}_3=(0,0.8)^{\tau},\qquad \boldsymbol{b}_4=(0,1)^{\tau}.$$

将模型(3.5.11)的样本轨道的二维序列$\{\boldsymbol{x}_t\}$绘在平面图上,它与一片蕨齿类的树叶很相似,如图 3.5.2 所示.

第 4 章

非线性时间序列模型的概率结构

本章主要讨论非线性时间序列模型的平稳性、遍历性以及高阶矩的存在性等问题. 讨论这些问题,需要借助第 2 章的离散动态系统的概念以及一般状态马氏链的理论和方法. 因此,在本章 §4.1 中作为预备知识,先介绍一般状态马氏链的有关概念和结论.

§4.1 一般状态马氏链

如第 3 章所述,任何高阶自回归模型都可表示成等价的一阶高维自回归模型. 因此高阶自回归模型的概率结构可借助于一阶高维自回归模型的概率结构进行研究. 对于一阶 m 维 LAR 模型

$$\boldsymbol{x}_t = \boldsymbol{\Phi}\boldsymbol{x}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{4.1.1}$$

其中$\{\boldsymbol{\varepsilon}_t\}$为 m 维向量白噪声序列,满足

$$\begin{cases} E\boldsymbol{\varepsilon}_t = 0, \quad E\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^{\tau} = \Gamma; \\ \boldsymbol{\varepsilon}_t \text{ 与 } \{\boldsymbol{x}_s, s < t\} \text{ 相互独立}. \end{cases} \tag{4.1.2}$$

为了讨论模型(4.1.1)的平稳解存在的条件,可以利用(4.1.1)式反复迭代而得到如下表达式:

$$\boldsymbol{x}_t = \boldsymbol{\varepsilon}_t + \boldsymbol{\Phi}\boldsymbol{\varepsilon}_{t-1} + \cdots + \boldsymbol{\Phi}^{n-1}\boldsymbol{\varepsilon}_{t-n+1} + \boldsymbol{\Phi}^n\boldsymbol{x}_{t-n}. \tag{4.1.3}$$

模型(4.1.1)能否有平稳解,与如下离散动态系统

$$\boldsymbol{y}_t = \boldsymbol{\Phi}\boldsymbol{y}_{t-1} \tag{4.1.4}$$

是否在 $\boldsymbol{y}=\boldsymbol{0}$ 为整体稳定有关. 由推论 2.3.10 知,线性系统(4.1.4)在 $\boldsymbol{y}=\boldsymbol{0}$ 处为整体稳定,当且仅当 $\boldsymbol{\Phi}$ 的特征值的模都小于 1. 利用 §2.1 中使用的约当标准形方法可知,如果 $\boldsymbol{\Phi}$ 的特征值 $\lambda_i(\boldsymbol{\Phi})$满足

$$\max_{1\leqslant i\leqslant m} |\lambda_i(\boldsymbol{\Phi})| < 1,$$

则随机级数

$$\sum_{j=0}^{\infty}\boldsymbol{\Phi}^j\boldsymbol{\varepsilon}_{t-j} \tag{4.1.5}$$

是均方收敛的. 经直接验证可知,由(4.1.5)式定义的随机序列

$$\boldsymbol{x}_t = \sum_{j=0}^{\infty} \boldsymbol{\Phi}^j \boldsymbol{\varepsilon}_{t-j}$$

满足 LAR 模型(4.1.1).

反之,若 $\boldsymbol{\Phi}$ 的特征值的模不全小于 1,且 $E\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^{\tau} = \boldsymbol{\Gamma}$ 为满秩矩阵,则(4.1.1)式一定没有平稳解. 这些都是文献中熟知的结果(参见文献[57]和[58]).

但是,对于 NLAR 模型

$$\boldsymbol{x}_t = \boldsymbol{\varphi}(\boldsymbol{x}_{t-1}) + \boldsymbol{\varepsilon}_t, \tag{4.1.6}$$

式中 $\boldsymbol{\varphi}$ 是 R^m 到 R^m 的可测函数,$\{\boldsymbol{\varepsilon}_t\}$是向量白噪声序列,满足(4.1.2)式,一般说来,模型(4.1.6)不像模型(4.1.1)那样,有形如(4.1.3)的明显解的表达式. 因此讨论模型(4.1.1)是否有平稳解的方法对模型(4.1.6)不再奏效. 但是应该注意到,模型(4.1.6)从初值 $\boldsymbol{x}_0$ 出发而得到迭代序列是一个马氏链,以$(R^m, \mathscr{B}_m)$为状态空间,其中 $\mathscr{B}_m$ 为 R^m 中一些子集所生成的 σ-代数,时间指标 t 取自然数. 这类马氏链属于近二十年来才被较多地研究的一般状态马氏链(见文献[83]、[114]及[115]). 利用一般状态马氏链的有关结果,可以解决许多 NLAR 模型的平稳性、遍历性和高阶矩的存在性等问题. 在本书中,起着重要作用的是一般状态马氏链的漂移准则以及判断平稳遍历性的定理. 所以本节不打算论及更多的一般状态马氏链的知识,而只是介绍一些必要的概念和结论,以助叙述漂移准则和判断定理. 所介绍的引理和定理由于要涉及过多的一般状态马氏链的知识而略去证明,对证明感兴趣的读者可参阅文献[83]、[114]和[115].

一、一般状态马氏链与转移概率

设$\{\boldsymbol{x}_n, n \geqslant 1\}$是概率空间$(\Omega, \mathscr{F}, P)$上取值于$(R^m, \mathscr{B}_m)$的一个随机向量序列,如果对任意的 $A \in \mathscr{B}_m$,都有

$$P(\boldsymbol{x}_t \in A | \boldsymbol{x}_{t-1}, \boldsymbol{x}_{t-2}, \cdots) = P(\boldsymbol{x}_t \in A | \boldsymbol{x}_{t-1}), \text{a.s.}, \tag{4.1.7}$$

则称$\{\boldsymbol{x}_n\}$是以$(R^m, \mathscr{B}_m)$为状态空间的马氏链.

条件(4.1.7)式称为马氏性,其直观涵义是:在给定全部历史值时,现在时刻的随机向量的条件分布只与前一时刻的 $\boldsymbol{x}_{t-1}$的取值有关.

在一般情况下,(4.1.7)式右端的条件概率依赖于集合 A、$\boldsymbol{x}_{t-1}$的取值和时间 t. 若它不依赖于时间 t,则称$\{\boldsymbol{x}_n\}$为齐时马氏链. 这时,对任意的 $\boldsymbol{x} \in R^m$ 和任意的 $A \in \mathscr{B}_m$,(4.1.7)式可表示为

$$P(\boldsymbol{x},A)=P(\boldsymbol{x}_t\in A\,|\,\boldsymbol{x}_{t-1}=\boldsymbol{x})$$
$$=P(\boldsymbol{x}_1\in A\,|\,\boldsymbol{x}_0=\boldsymbol{x}). \tag{4.1.8}$$

式中 P 称为齐时马氏链$\{\boldsymbol{x}_n\}$的转移概率.(4.1.8)式表明了对任意给定的 $A\in\mathscr{B}_m$,$P(\cdot,A)$是$(R^m,\mathscr{B}_m)$上的一个可测函数,而对给定的 $\boldsymbol{x}\in R^m$,$P(\boldsymbol{x},\cdot)$是$(R^m,\mathscr{B}_m)$上的概率测度.

对于齐时马氏链,从 $\boldsymbol{x}_0=\boldsymbol{x}$ 转移到 $\boldsymbol{x}_2\in A$ 的条件概率为

$$P_2(\boldsymbol{x},A)=P(\boldsymbol{x}_2\in A\,|\,\boldsymbol{x}_0=\boldsymbol{x})$$
$$=\int_{R^m}P(\boldsymbol{y},A)P(\boldsymbol{x},\mathrm{d}\boldsymbol{y}).$$

类似地,从 $\boldsymbol{x}_0=\boldsymbol{x}$ 出发,经 n 步转移到 $\boldsymbol{x}_n\in A$ 的条件概率为

$$P_n(\boldsymbol{x},A)=P(\boldsymbol{x}_n\in A\,|\,\boldsymbol{x}_0=\boldsymbol{x})$$
$$=\int_{R^m}P_{n-1}(\boldsymbol{y},A)P(\boldsymbol{x},\mathrm{d}\boldsymbol{y}),\quad n\geqslant 1. \tag{4.1.9}$$

式中 $P_n(\boldsymbol{x},A)$称为齐时马氏链$\{\boldsymbol{x}_n\}$的 *n 步转移概率*.

以下无特别声明的,所讨论的马氏链均为齐时马氏链.在下面的论述中略去"齐时"二字.

下面讨论马氏链$\{\boldsymbol{x}_n\}$的有限维联合分布的表示式.设 $\boldsymbol{x}_0$ 的概率分布 P_0 为$\{\boldsymbol{x}_n\}$的初始分布,则由马氏链和条件概率的性质,对任意的 $A\in\mathscr{B}_m$,有

$$P(\boldsymbol{x}_1\in A)=\int_{R^m}P(\boldsymbol{y},A)P_0(\mathrm{d}\boldsymbol{y})$$
$$=\int_{R^m}P_0(\mathrm{d}\boldsymbol{y})\int_A P(\boldsymbol{y},\mathrm{d}\boldsymbol{x}).$$

一般地,有

$$\begin{aligned}P(\boldsymbol{x}_n\in A)&=\int_{R^m}P_n(\boldsymbol{y},A)P_0(\mathrm{d}\boldsymbol{y})\\&=\int_{R^m}P_0(\mathrm{d}\boldsymbol{y})\int_A P_n(\boldsymbol{y},\mathrm{d}\boldsymbol{x}).\end{aligned} \tag{4.1.10}$$

由马氏链和条件概率的性质,对任意的 A_0 和 $A_1\in\mathscr{B}_m$,有

$$P(\boldsymbol{x}_0\in A_0,\boldsymbol{x}_1\in A_1)=\int_{A_0}P(\boldsymbol{x}_0,A_1)P_0(\mathrm{d}\boldsymbol{x}_0)$$
$$=\int_{A_0}P_0(\mathrm{d}\boldsymbol{x}_0)\int_{A_1}P(\boldsymbol{x}_0,\mathrm{d}\boldsymbol{x}_1).$$

一般地,对任意的 $A_0,A_1,\cdots,A_N\in\mathscr{B}_m$,有

$$P(\boldsymbol{x}_0\in A_0,\boldsymbol{x}_1\in A_1,\cdots,\boldsymbol{x}_N\in A_N)$$

$$= \int_{A_0} P_0(\mathrm{d}\boldsymbol{x}_0) \int_{A_1} P(\boldsymbol{x}_0, \mathrm{d}\boldsymbol{x}_1) \cdots \int_{A_N} P(\boldsymbol{x}_{n-1}, \mathrm{d}\boldsymbol{x}_n). \quad (4.1.11)$$

特别地，当(4.1.11)式中的所有分布都有密度时，(4.1.11)式成为

$$P(\boldsymbol{x}_0 \in A_0,\ \boldsymbol{x}_1 \in A_1, \cdots, \boldsymbol{x}_N \in A_N)$$
$$= \int_{A_0} \cdots \int_{A_N} f(\boldsymbol{y}_N | \boldsymbol{y}_{N-1}) \cdot \cdots \cdot f(\boldsymbol{y}_1 | \boldsymbol{y}_0) f(\boldsymbol{y}_0) \mathrm{d}\boldsymbol{y}_N \cdots \mathrm{d}\boldsymbol{y}_0,$$

式中 $f(\boldsymbol{y}_i | \boldsymbol{y}_{i-1})$ 表示在给定 $\boldsymbol{x}_{i-1} = \boldsymbol{y}_{i-1}$ 的条件下 $\boldsymbol{x}_i$ 的条件密度，$f(\boldsymbol{y}_0)$ 为 $\boldsymbol{x}_0$ 的密度. 记 $f(\boldsymbol{y}_0, \boldsymbol{y}_1, \cdots, \boldsymbol{y}_N)$ 为 $(\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_N)^\tau$ 的联合密度，则上式意味着

$$f(\boldsymbol{y}_0, \boldsymbol{y}_1, \cdots, \boldsymbol{y}_N) = f(\boldsymbol{y}_N | \boldsymbol{y}_{N-1}) \cdot \cdots \cdot f(\boldsymbol{y}_1 | \boldsymbol{y}_0) f(\boldsymbol{y}_0).$$

若 $0 \leqslant k_1 < k_2 < \cdots < k_s \leqslant N$，则由(4.1.11)式不难得到 $(\boldsymbol{x}_{k_1}, \cdots, \boldsymbol{x}_{k_s})^\tau$ 的边缘分布. 可见马氏链 $\{\boldsymbol{x}_n\}$ 的任何有限维分布均被其转移概率和初始分布所唯一决定. 按 § 1.1 所述，其概率结构亦被转移概率和初始分布所唯一决定.

如果马氏链 $\{\boldsymbol{x}_n\}$ 的某个初始分布 P_0 满足如下性质：

对任意的 $A \in \mathscr{B}_m$，有

$$P(\boldsymbol{x}_1 \in A) = \int_{R^m} P(\boldsymbol{y}, A) P_0(\mathrm{d}\boldsymbol{y}) = P_0(\boldsymbol{x}_0 \in A),$$

则由(4.1.9)、(4.1.11)和上式，有

$$\begin{aligned} P(\boldsymbol{x}_2 \in A) &= \int_{R^m} P_2(\boldsymbol{y}, A) P_0(\mathrm{d}\boldsymbol{y}) \\ &= \iint_{R^m} P(\boldsymbol{z}, A) P(\boldsymbol{y}, \mathrm{d}\boldsymbol{z}) P_0(\mathrm{d}\boldsymbol{y}) \\ &= \int_{R^m} P(\boldsymbol{z}, A) P_0(\mathrm{d}\boldsymbol{z}) \\ &= P_0(\boldsymbol{x}_0 \in A). \end{aligned}$$

依此递推可知，对任意的 $A \in \mathscr{B}_m$ 和 $n \geqslant 1$，有

$$P(\boldsymbol{x}_n \in A) = P_0(\boldsymbol{x}_0 \in A), \quad (4.1.12)$$

这时 P_0 称为马氏链 $\{\boldsymbol{x}_n\}$ 的**不变分布**. (4.1.12)式表明从不变分布 P_0 出发，通过转移概率构造的马氏链是平稳的. 但是需要指出的是：一个马氏链既不一定总有不变分布，也不一定只有一个不变分布.

二、不可约、非周期和小集的概念

定义 4.1.1 设 $\{\boldsymbol{x}_n\}$ 为马氏链，如果存在 $(R^m, \mathscr{B}_m)$ 上的测度 ν，使得对每个 $A \in \mathscr{B}_m$，且 $\nu(A) > 0$，以及任意的 $\boldsymbol{x} \in R^m$，都有

$$\sum_{n=1}^{\infty} P_n(\boldsymbol{x}, A) > 0, \tag{4.1.13}$$

则称$\{\boldsymbol{x}_n\}$为 ν-不可约的.

(4.1.13)式表明：马氏链$\{\boldsymbol{x}_n\}$从任何初始值 $\boldsymbol{x}_0=\boldsymbol{x}$ 出发，在有限步之内到达 ν-正测集 A 的概率总是正的.

如果测度 ν' 相对于测度 ν 是绝对连续的，即对于 $A\in\mathscr{B}_m$，当 $\nu'(A)>0$ 时，必有 $\nu(A)>0$，则一个 ν-不可约马氏链必为 ν'-不可约的. 由此可见，不可约马氏链的不可约测度并不唯一. 为了使之唯一化，我们引入最大不可约测度 ν^* 如下，即

(i) $\{\boldsymbol{x}_n\}$是 ν^*-不可约的；

(ii) 如果$\{\boldsymbol{x}_n\}$是 ν-不可约的，则 ν 必是相对于 ν^* 绝对连续的；

(iii) $\nu^*(A)=0\Rightarrow\nu^*\left(\left\{\boldsymbol{x}: \sum_{n=1}^{\infty} P_n(\boldsymbol{x}, A)>0\right\}\right)=0.$

以后我们说马氏链$\{\boldsymbol{x}_n\}$是不可约的，总是指相对于最大不可约测度而言的. 与不可约相对立的概念是可约，下面举一例子说明之.

例 4.1.2　考虑如下随机模型：

$$x_t=\frac{1}{2}x_{t-1}+\left\{\frac{1}{2}-I(x_{t-1}\leqslant 0)\right\}\varepsilon_t,\ t\geqslant 1, \tag{4.1.14}$$

式中$\{\varepsilon_t\}$为独立同分布序列，且 ε_t 服从[1,2]上的均匀分布. 易见，满足(4.1.14)式的随机序列是齐时马氏链. 但是当 $x_0=x<0$，$A=(a,\infty)$，$a>0$时，有

$$P(x,A)=P\left(\left(\frac{1}{2}x_0-\frac{1}{2}\varepsilon_1\right)\in A\,|\,x_0=x\right)$$

$$=P\left(\frac{1}{2}(x-\varepsilon_1)>a\right)=0.$$

依次类推可知，对任意的 $n\geqslant 1$，有

$$P_n(x,A)=0.$$

类似地，若取 $x_0=y\geqslant 0$，$B=(-\infty,b)$，且 $b<0$，则有

$$P_n(y,B)=0\quad(n\geqslant 1).$$

不难看出，满足(4.1.14)式的马氏链不是不可约的. 实际上(4.1.14)式可分成两个不同的随机模型：

$$x_t=\frac{1}{2}x_{t-1}+\frac{1}{2}\varepsilon_t,\quad 当\ x_0>0;$$

$$x_t = \frac{1}{2}x_{t-1} - \frac{1}{2}\varepsilon_t, \quad 当\ x_0 \leqslant 0.$$

对于以上两个模型可以分别进行研究. 因此以下主要讨论不可约马氏链.

在介绍马氏链的非周期概念之前,先给出马氏链$\{\boldsymbol{x}_n\}$的循环集的定义.

定义 4.1.3 设 $A_1, A_2, \cdots, A_d$ 为 $\mathscr{B}_m$ 中一组互不相交的集合,称为马氏链$\{\boldsymbol{x}_n\}$的一组长度为 d 的循环集,如果

$$P(\boldsymbol{x}, A_j) = 1, \ \boldsymbol{x} \in A_{j-1} \quad (j = 2, \cdots, d);$$
$$P(\boldsymbol{x}, A_1) = 1, \ \boldsymbol{x} \in A_d.$$

定义 4.1.4 设$\{\boldsymbol{x}_n\}$为马氏链, $A_1, A_2, \cdots, A_d$ 是它的循环集,且满足如下两个条件:

(i) 存在$(R^m, \mathscr{B}_m)$上的非平凡测度 ν,使得 $\nu(A_i) > 0 \quad (i = 1, \cdots, d)$,且

$$\nu\left(R^m \setminus \bigcup_{i=1}^{d} A_i\right) = 0,$$

其中记号"$A \setminus B$"表示集合 A 与 B 的差集;

(ii) 如果 $\widetilde{A}_1, \widetilde{A}_2, \cdots, \widetilde{A}_{d'}$ 为$\{\boldsymbol{x}_n\}$的另一组循环集,则有 $d' | d$.

当 $d=1$ 时,$\{\boldsymbol{x}_n\}$称为非周期马氏链;当 $d>1$ 时,$\{\boldsymbol{x}_n\}$称为周期马氏链,周期为 d.

对于具有周期 d 的周期马氏链$\{\boldsymbol{x}_n\}$,设 $A_1, \cdots, A_d$ 为满足上述定义 4.1.4 条件的循环集. 若 $\boldsymbol{x}_0 \in A_1$,则由循环集的定义知

$$\begin{aligned} P(\boldsymbol{x}_1 \in A_2) &= P(\boldsymbol{x}_2 \in A_3) = \cdots \\ &= P(\boldsymbol{x}_{d-1} \in A_d) = P(\boldsymbol{x}_d \in A_1) = 1. \end{aligned}$$

由此类推,有

$$P(\boldsymbol{x}_{td} \in A_1) = 1, \ t \geqslant 1.$$

定义随机序列 $\boldsymbol{y}_t = \boldsymbol{x}_{td}, t \geqslant 1$,则$\{\boldsymbol{y}_t\}$是一个以$(A_1, A_1 \cap \mathscr{B}_m)$为状态空间的非周期马氏链. 同样,若 $\boldsymbol{x}_0 \in A_j \quad (2 \leqslant j \leqslant d)$,可得另一个以$(A_j, A_j \cap \mathscr{B}_m)$为状态空间的非周期不可约马氏链. 这样可得到 d 个分别以$(A_j, A_j \cap \mathscr{B}_m) \ (j = 1, \cdots, d)$为状态空间的非周期马氏链,可分别讨论之.

例 4.1.5 考虑如下分段线性自回归模型:

$$x_t = \begin{cases} -\dfrac{1}{2}x_{t-1} + \varepsilon_t, & 当\ x_{t-1} < 0; \\ -\dfrac{1}{2}x_{t-1} - \varepsilon_t, & 当\ x_{t-1} \geqslant 0, \end{cases} \tag{4.1.15}$$

式中$\{\varepsilon_t\}$为独立同分布序列，且ε_t服从(1,2)上的均匀分布. 取$A_1=(-\infty,0)$，$A_2=[0,\infty)$，则容易验证

$$P(x,A_2)=1,\ \forall\ x\in A_1;$$
$$P(x,A_1)=1,\ \forall\ x\in A_2.$$

可见，由模型(4.1.15)产生的随机序列是周期马氏链，周期为2. 将模型(4.1.15)写为

$$x_t=\varphi(x_{t-1},\varepsilon_t)=\begin{cases}-\dfrac{1}{2}x_{t-1}+\varepsilon_t, & \text{当 } x_{t-1}<0;\\ -\dfrac{1}{2}x_{t-1}-\varepsilon_t, & \text{当 } x_{t-1}\geqslant 0.\end{cases}$$

依此可得

$$x_1=\varphi(x_0,\varepsilon_1)=\begin{cases}-\dfrac{1}{2}x_0+\varepsilon_1, & \text{当 } x_0<0;\\ -\dfrac{1}{2}x_0-\varepsilon_1, & \text{当 } x_0\geqslant 0.\end{cases}$$

由此可见，当$x_0<0$时，必有$x_1>0$；当$x_0\geqslant 0$时，必有$x_1<0$. 于是又有

$$x_2=\varphi(x_1,\varepsilon_2)=\begin{cases}-\dfrac{1}{2}x_1+\varepsilon_2, & \text{当 } x_1<0;\\ -\dfrac{1}{2}x_1-\varepsilon_2, & \text{当 } x_1\geqslant 0,\end{cases}$$

$$=\begin{cases}\dfrac{1}{4}x_0+\dfrac{1}{2}\varepsilon_1+\varepsilon_2, & \text{当 } x_0\geqslant 0;\\ \dfrac{1}{4}x_0-\dfrac{1}{2}\varepsilon_1-\varepsilon_2, & \text{当 } x_0<0.\end{cases}$$

类似地，有

$$x_3=\varphi(x_2,\varepsilon_3)=\begin{cases}-\dfrac{1}{2}x_2+\varepsilon_3, & \text{当 } x_2<0;\\ -\dfrac{1}{2}x_2-\varepsilon_3, & \text{当 } x_2\geqslant 0,\end{cases}$$

$$=\begin{cases}\dfrac{1}{4}x_1+\dfrac{1}{2}\varepsilon_2+\varepsilon_3, & \text{当 } x_1\geqslant 0;\\ \dfrac{1}{4}x_1-\dfrac{1}{2}\varepsilon_2-\varepsilon_3, & \text{当 } x_1<0.\end{cases}$$

一般地，对$t=1,2,\cdots$，有如下递推关系式：

$$x_{2t}=\begin{cases}\frac{1}{4}x_{2t-2}+\frac{1}{2}\varepsilon_{2t-1}+\varepsilon_{2t}, & \text{当 } x_{2t-2}\geqslant 0;\\ \frac{1}{4}x_{2t-2}-\frac{1}{2}\varepsilon_{2t-1}-\varepsilon_{2t}, & \text{当 } x_{2t-2}<0.\end{cases}$$

$$x_{2t+1}=\begin{cases}\frac{1}{4}x_{2t-1}+\frac{1}{2}\varepsilon_{2t}+\varepsilon_{2t+1}, & \text{当 } x_{2t-1}\geqslant 0;\\ \frac{1}{4}x_{2t-1}-\frac{1}{2}\varepsilon_{2t}-\varepsilon_{2t+1}, & \text{当 } x_{2t-1}<0.\end{cases}$$

如果记 $y_t=x_{2t}$，$z_t=x_{2t+1}$，则容易验证 $\{y_t\}$ 和 $\{z_t\}$ 都是非周期马氏链. 因此，对具有周期为 2 的马氏链(4.1.15)只需分别研究 $\{y_t\}$ 和 $\{z_t\}$ 这两个非周期马氏链即可. 本章下文只考虑非周期马氏链.

在讨论由非线性时间序列模型所决定的马氏链的遍历性时. 常需要判别其非周期性. 下面的引理是很有用的.

引理 4.1.6　设 $\{x_n\}$ 是 ν-不可约的马氏链，它是非周期的，当且仅当存在 $A\in\mathscr{B}_m$，且 $\nu(A)>0$，对 A 的每个 ν-正测子集 $B\subset A$，$\nu(B)>0$，都存在整数 n，使得对任何 $\boldsymbol{x}\in B$，有

$$P_n(\boldsymbol{x},B)>0,\quad P_{n+1}(\boldsymbol{x},B)>0. \tag{4.1.16}$$

引理 4.1.6 的证明见文献[106].

由前面的讨论知，一个马氏链的不变概率分布是它的一个平稳分布. 下文将要讨论的遍历性刻划了马氏链的转移概率 $P_n(\boldsymbol{x},\cdot)$ 向其平稳分布的收敛性态. 为了研究非线性时间序列模型是否有平稳解，此解是否为遍历等问题，一个重要的途径是借助于研究马氏链的遍历性来解决. 在研究马氏链的遍历性时，如下的小集概念起着重要的作用，但是这一概念却不易掌握. 为了有助于理解小集的涵义，先回顾一下取值为可列状态的马氏链的遍历性条件. 不妨考虑在 $\{1,2,\cdots\}$ 中取值的马氏链 $\{x_n\}$，将其转移概率记为

$$p_{ij}=P(x_n=j\mid x_{n-1}=i)\quad(i,j=1,2,\cdots). \tag{4.1.17}$$

如果此马氏链是不可约的、非周期的，且存在有限子集 $C\subset\{1,2,\cdots\}$、非负函数 g 及正常数 K 和 δ，使得

(i) $E\{g(x_{n+1})\mid x_n=j\}-g(j)\leqslant K$，当 $j\in C$；

(ii) $E\{g(x_{n+1})\mid x_n=j\}-g(j)\leqslant -\delta$，当 $j\overline{\in} C$，

则满足(4.1.17)式的马氏链为遍历的，且

$$\lim_{n\to\infty}p_{ij}^{(n)}=\lim_{n\to\infty}P(x_n=j\mid x_0=i)=\pi_j$$
$$(j=1,2,\cdots),$$

其中 π_j 与 i 无关，且

$$\sum_{j=1}^{\infty}\pi_j = 1.$$

换言之，$P(x_0=j)=\pi_j$ $(j=1,2,\cdots)$是唯一的平稳分布.

在考虑一般状态马氏链时，为了获得与上述结果相应的结论，需要如下的小集概念，它是上述有限点集 C 的推广定义. 由于本书不需要论及太多专门的一般状态马氏链的概念，故在此仅根据本书的需要，先给出小集的定义，然后介绍两个判断小集的方法.

定义 4.1.7 一个非空集 $C\in\mathscr{B}_m$ 称为小集(Small set)，如果存在一个正整数 k、正常数 b 和概率测度 ν，使得对任意的 $A\in\mathscr{B}_m$ 和 $\boldsymbol{x}\in C$，有

$$P_k(\boldsymbol{x},A)\geqslant b\nu(A). \qquad (4.1.18)$$

引理 4.1.8 设$\{\boldsymbol{x}_n\}$是 ν-不可约的非周期马氏链，对于 $C\in\mathscr{B}_m$，$\nu(C)>0$，如果存在 $A\in\mathscr{B}_m$，$\nu(A)>0$ 和正整数 l，使得对任意的 $B\subset A$，$\nu(B)>0$，都有

$$\inf_{\boldsymbol{x}\in C}\sum_{n=1}^{l}P_n(\boldsymbol{x},B)>0, \qquad (4.1.19)$$

则 C 是一个小集.

引理的证明从略(参见文献[83]).

引理 4.1.9 设$\{\boldsymbol{x}_n\}$是弱连续的马氏链，即对一切有界连续函数 $g:R^m\to R^1$，$\int g(\boldsymbol{y})P(\boldsymbol{x},\mathrm{d}\boldsymbol{y})$关于 $\boldsymbol{x}$ 是连续的，则一切 ν 非零测度的相对紧集(relatively compact set)是小集.

引理的证明见文献[106].

三、遍历性与几何遍历性

以 π 记马氏链$\{\boldsymbol{x}_n\}$的不变概率分布，则对 $A\in\mathscr{B}_m$，有

$$P(\boldsymbol{x}_0\in A)=P(\boldsymbol{x}_1\in A)=\int P(\boldsymbol{x},A)\pi(\mathrm{d}\boldsymbol{x}). \qquad (4.1.20)$$

定义 4.1.10 (i) 马氏链$\{\boldsymbol{x}_n\}$称为遍历的，如果存在一概率测度 π，使得对任意的 $\boldsymbol{x}\in R^m$，有

$$\lim_{n\to\infty}\|P_n(\boldsymbol{x},\cdot)-\pi(\cdot)\|_\tau=0, \qquad (4.1.21)$$

其中 $\|\cdot\|_\tau$ 表示全变差范数.

(ii) 如果还存在常数 $0\leqslant\rho<1$，使得对任意的 $\boldsymbol{x}\in R^m$，有

$$\lim_{n\to\infty}\rho^n\|P_n(\boldsymbol{x},\cdot)-\pi(\cdot)\|_\tau=0, \qquad (4.1.22)$$

则称$\{\boldsymbol{x}_n\}$为几何遍历的.

注意:这里的全变差范数是指对广义测度而言的.设ν为广义测度,$|\nu|=\nu^+ +\nu^-$,其中

$$\nu^+(A)=\int_A\{\nu(\mathrm{d}\boldsymbol{x})\vee 0\},\quad A\in\mathscr{B}_m,$$

$$-\nu^-(A)=\int_A\{\nu(\mathrm{d}\boldsymbol{x})\wedge 0\},\quad A\in\mathscr{B}_m,$$

式中$a\vee b=\max\{a,b\}$, $a\wedge b=\min\{a,b\}$,则

$$\|\nu\|_\tau=\int_{R^m}|\nu(\mathrm{d}\boldsymbol{x})|$$

$$=\int_{R^m}\nu^+(\mathrm{d}\boldsymbol{x})+\int_{R^m}\nu^-(\mathrm{d}\boldsymbol{x}).$$

有了以上的准备之后,下面叙述两个重要的判别马氏链的遍历性的定理.

定理 4.1.11　设$\{\boldsymbol{x}_n\}$为不可约的、非周期的马氏链,如果存在一个非负可测函数g,一个小集C和常数$c_1>0$, $c_2>0$,使得

(i) $E\{g(\boldsymbol{x}_n)|\boldsymbol{x}_{n-1}=\boldsymbol{x}\}\leqslant g(\boldsymbol{x})-c_1,\quad\forall\,\boldsymbol{x}\overline{\in}C;$　　(4.1.23)

(ii) $E\{g(\boldsymbol{x}_n)|\boldsymbol{x}_{n-1}=\boldsymbol{x}\}\leqslant c_2,\quad\forall\,\boldsymbol{x}\in C,$　　(4.1.24)

则$\{\boldsymbol{x}_n\}$为遍历的.

定理 4.1.12　设$\{\boldsymbol{x}_n\}$为不可约的、非周期的马氏链.如果存在一个非负可测函数g,一个小集C和常数$c_1>0$, $c_2>0$, $0<\rho<1$,使得

(i) $E\{g(\boldsymbol{x}_n)|\boldsymbol{x}_{n-1}=\boldsymbol{x}\}\leqslant\rho g(\boldsymbol{x})-c_1,\quad\forall\,\boldsymbol{x}\overline{\in}C;$　　(4.1.25)

(ii) $E\{g(\boldsymbol{x}_n)|\boldsymbol{x}_{n-1}=\boldsymbol{x}\}\leqslant c_2,\quad\forall\,\boldsymbol{x}\in C,$　　(4.1.26)

则$\{\boldsymbol{x}_n\}$为几何遍历的.

在定理 4.1.11 和定理 4.1.12 中,非负函数g称为*准则函数*.两个定理分别称为*遍历性漂移准则*和*几何遍历性漂移准则*.定理的证明见文献[83]或[114].

定理 4.1.13　设$\{\boldsymbol{x}_n\}$为不可约的、非周期的马氏链.$\{\boldsymbol{x}_n\}$为遍历(或几何遍历)的,当且仅当存在一个正整数$l\geqslant1$,使得$\{\boldsymbol{x}_{ln},n\geqslant0\}$为遍历(或几何遍历)的.

定理 4.1.13 的证明见文献[102].

本章主要讨论几何遍历性,所以定理 4.1.12 和定理 4.1.13 起着重要的作用.除以上所介绍的与本书有直接关系的知识外,一般状态马氏链还有很丰富的内容,但这些不属于本书的讨论范围.

§4.2 非线性自回归模型的几何遍历性

本节借助于§4.1所介绍的一般状态马氏链的遍历性定理来讨论可加噪声NLAR模型和函数型随机条件方差AR模型的几何遍历性. 为此,先介绍NLAR模型的平稳分布、遍历性和几何遍历性等概念. 然后讨论由NLAR模型所决定的马氏链为不可约、非周期的条件.

一、NLAR模型的平稳性与遍历性

考虑如下一般的一阶 m 维NLAR模型

$$\begin{cases} \boldsymbol{x}_t = \boldsymbol{\varphi}(\boldsymbol{x}_{t-1}, \boldsymbol{\varepsilon}_t), t \geqslant 1; \\ \boldsymbol{x}_0 \sim F_0, \end{cases} \tag{4.2.1}$$

式中 $\{\boldsymbol{\varepsilon}_t\}$ 为 l 维向量白噪声序列,且 ε_t 与 $\boldsymbol{x}_0$ 独立, $\boldsymbol{x}_0$ 是 m 维初始随机向量,其概率分布为 F_0. $\boldsymbol{\varphi}$ 是 R^{m+l} 到 R^m 的可测函数. 显然,可加噪声NLAR模型(3.5.8)和函数型随机条件方差NLAR模型(3.5.7)都是(4.2.1)式的特例. 在模型(3.3.22)中,令 $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \cdots, \varepsilon_{lt})^{\tau}$,若

$$\varphi(\boldsymbol{x}_{t-1}, \boldsymbol{\varepsilon}_t) = \left(\sum_{j=1}^{l}\{\varphi_{j0} + \sum_{k=1}^{p}\varphi_{jk}x_{t-j} + \varepsilon_{jt}\}I(x_{t-d} \in R_j), x_{t-1}, \cdots, x_{t-p+1}\right)^{\tau},$$

则模型(3.3.22)也可化为模型(4.2.1)的特殊情况. 对 β-ARCH模型(3.3.14)和(3.3.15)如果 $\{\varepsilon_t\}$ 满足条件(3.2.2)式,若记 $y_t = \log e_t^2$, $\eta_t = \log\varepsilon_t^2$,对(3.3.14)式中第二个等式两端平方,并取对数,由(3.3.15)式得

$$y_t = \log(\alpha_0 + \alpha_1 e^{\beta y_{t-1}} + \cdots + \alpha_p e^{\beta y_{t-p}}) + \eta_t. \tag{4.2.2}$$

显然,(4.2.2)可化为可加噪声NLAR模型(3.5.8)的特例,从而 β-ARCH模型也是模型(4.2.1)的特例. 但是§3.5中的双重线性模型一般不属于(4.2.1)式的NLAR模型类. 尽管如此,对模型(4.2.1)的解引入平稳性、遍历性和几何遍历性等概念,以及讨论其与齐时马氏链的联系仍有普遍的意义.

定义4.2.1 设 m 维向量随机序列 $\{\boldsymbol{x}_t\}$ 服从一阶 m 维NLAR模型(4.2.1):

(i) 如果当 $\boldsymbol{x}_0 \sim F$ 时, $\boldsymbol{x}_1 = \boldsymbol{\varphi}(\boldsymbol{x}_0, \boldsymbol{\varepsilon}_1) \sim F$,则称 F 为模型(4.2.1)的不变概率分布;

(ii) 如果 $\boldsymbol{x}_0 \sim F$,且 F 为模型(4.2.1)的不变概率分布,则称从 $\boldsymbol{x}_0$ 出

发由(4.2.1)式产生的迭代序列$\{x_t, t\geqslant 0\}$为模型(4.2.1)的平稳解.

需要注意的是:一个时间序列模型既不一定总有不变概率分布,也不一定只有一个不变概率分布.下面举两个例子说明之.

例 4.2.2 考虑如下一阶自回归模型:

$$x_t = x_{t-1} + \varepsilon_t, \tag{4.2.3}$$

其中$\{\varepsilon_t\}$为白噪声序列,ε_t 与 x_{t-1}独立.

如果 ε_t 为非退化的,无论 $E\varepsilon_t^2$ 是否为有穷的,模型(4.2.3)总没有不变概率分布.

事实上,如果(4.2.3)式有不变概率分布,记其相应的特征函数为 f_x, ε_t 的特征函数记为 f_ε,由 ε_t 与 x_{t-1}独立的假定及(4.2.3)式,有

$$f_x(s) = f_x(s) f_\varepsilon(s), \quad \forall\, s \in R^1.$$

由上式可得

$$|f_x(s)| = |f_x(s)|\,|f_\varepsilon(s)|, \quad \forall\, s \in R^1. \tag{4.2.4}$$

由于 ε_t 是非退化的,故存在 $\delta_1 > 0$,使得

$$|f_\varepsilon(s)| < 1, \quad \forall\, s \in (0, \delta_1). \tag{4.2.5}$$

另外,模型(4.2.3)的分布显然是非退化的,从而存在 $\delta_2 > 0$,使得

$$|f_x(s)| < 1, \quad \forall\, s \in (0, \delta_2).$$

取 $\delta = \min(\delta_1, \delta_2)$,则由(4.2.4)式有

$$|f_\varepsilon(s)| \equiv 1, \quad \forall\, s \in (0, \delta).$$

这与(4.2.5)式矛盾.故模型(4.2.3)没有不变概率分布,从而没有平稳解.

例 4.2.3 考虑如下分段线性 AR 模型:

$$x_t = \begin{cases} \dfrac{1}{2} + \dfrac{1}{2}x_{t-1} + \varepsilon_t, & \text{当 } x_{t-1} \geqslant 0; \\ -\dfrac{1}{2} + \dfrac{1}{2}x_{t-1} + \varepsilon_t, & \text{当 } x_{t-1} < 0, \end{cases} \tag{4.2.6}$$

其中$\{\varepsilon_t\}$为白噪声序列,$P\{\varepsilon_t = \frac{1}{4}\} = P\{\varepsilon_t = -\frac{1}{4}\} = \frac{1}{2}$,由直接验证知

$$u_t = 1 + \sum_{j=0}^{\infty} \left(\frac{1}{2}\right)^j \varepsilon_{t-j},$$

$$v_t = -1 + \sum_{j=0}^{\infty} \left(\frac{1}{2}\right)^j \varepsilon_{t-j}$$

都是模型(4.2.6)的解,从而 u_0 和 v_0 的分布都是模型(4.2.6)的不变概率分布,亦即说明模型(4.2.6)的不变概率分布不是唯一的.

这样,自然需要讨论怎样的 NLAR 模型(4.2.1)才有不变概率分布?其不变概率分布在什么条件下才是唯一的?如果模型(4.2.1)没有不变概率分布,显然它没有平稳解.如果模型(4.2.1)有不变概率分布,但不唯一,于是模型(4.2.1)有多个平稳解.这些都不属于本书所研究的对象.我们感兴趣的是模型(4.2.1)有唯一的不变概率分布的情况.对此情况,还要讨论从任何初始状态出发,$\boldsymbol{x}_t$ 的概率分布向其不变概率分布收敛的特性.下面的定义就是用来表述这些特性的.

定义 4.2.4 设模型(4.2.1)有唯一的不变概率分布 F,且对任何初始状态 $\boldsymbol{x}_0=\boldsymbol{x}$,即初始概率分布 $F_0^x=\delta_x$(这里 δ_x表示在 $\boldsymbol{x}$ 处的退化分布),由(4.2.1)式迭代产生的 $\boldsymbol{x}_t$ 的概率分布记为 F_t^x.如果

$$\lim_{t\to\infty}\| F_t^x - F \|_{\tau} = 0, \tag{4.2.7}$$

则称模型(4.2.1)为*遍历的*.

进一步,如果还存在正常数 $\rho<1$,使得

$$\lim_{t\to\infty}\rho^{-t}\| F_t^x - F \|_{\tau} = 0, \tag{4.2.8}$$

则称模型(4.2.1)为*几何遍历的*.

不变概率分布对模型(4.2.1)具有重要意义.从统计分析角度来看,上述的遍历性是进行统计分析的重要依据.也就是说,由模型(4.2.1)产生的可观测序列 $\boldsymbol{x}_t$ 的分布 F_t 是未知的,且与初始值有关.但是如果它满足(4.2.7)式或(4.2.8)式,便可提供统计估计未知的不变概率分布 F 的可能性.

二、NLAR 模型与一般状态马氏链的联系

为了研究模型(4.2.1)的几何遍历性,需要借助于一般状态马氏链的工具.如何将模型(4.2.1)与一般状态马氏链相联系,这是本小节要讨论的内容.

对于给定的模型(4.2.1),可定义如下的一步转移概率:

$$\begin{aligned} &P(\boldsymbol{x}_t \in A \mid \boldsymbol{x}_{t-1} = \boldsymbol{x}) \\ &= P(\boldsymbol{\varphi}(\boldsymbol{x}_{t-1},\varepsilon_t) \in A \mid \boldsymbol{x}_{t-1} = \boldsymbol{x}) \\ &= P(\boldsymbol{\varphi}(\boldsymbol{x},\varepsilon_t) \in A), \quad \forall\, \boldsymbol{x} \in R^m,\ A \in \mathscr{B}_m. \end{aligned} \tag{4.2.9}$$

由于$\{\varepsilon_t\}$是平稳的,(4.2.9)式右端的概率只与 A 和 $\boldsymbol{x}$ 有关,而与 t 无关.这表明由(4.2.9)式决定了一个形如(4.1.8)式的齐时马氏链的转移概率.从而由(4.2.1)式决定的$\{\boldsymbol{x}_t,t\geqslant 0\}$便是一个以$(R^m,\mathscr{B}_m)$为状态空间

的齐时马氏链. 研究模型(4.2.1)的几何遍历性问题便可转化为研究相应的马氏链的几何遍历性问题.

如上所述,模型(4.2.1)可确定一马氏链. 确切地说,就是由模型(4.2.1)和(4.2.9)式确定了齐时马氏链的转移概率. 但是当模型(4.2.1)的变换 φ 已知时,即使 ε_t 的分布类型给定了,而用(4.2.9)式得到其转移概率的明显表达式,也是十分困难的.

首先考虑 $m=1$ 的 LAR 模型

$$x_t = \alpha x_{t-1} + \varepsilon_t,\ t \geqslant 1, \tag{4.2.10}$$

式中$\{\varepsilon_t\}$为白噪声序列,满足(3.2.2)式,且参数 α 满足$|\alpha|<1$.

当$\{\varepsilon_t\}$为正态白噪声序列时,即 $\varepsilon_t \sim N(0,\sigma^2)$,则由(4.2.9)式知

$$\begin{aligned} P(x,(-\infty,a)) &= P(x_1 < a \mid x_0 = x) \\ &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{a-\alpha x} e^{-\frac{y^2}{2\sigma^2}}\,\mathrm{d}y \\ &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{a} e^{-\frac{(z-\alpha x)^2}{2\sigma^2}}\,\mathrm{d}z. \end{aligned}$$

上式表明了在给定 $x_{t-1}=x$ 的条件下,x_t 的转移概率为正态 $N(\alpha x,\sigma^2)$分布.

由(4.2.10)式递推,可得

$$x_n = \varepsilon_n + \alpha\varepsilon_{n-1} + \cdots + \alpha^{n-1}\varepsilon_1 + \alpha^n x_0.$$

由于

$$\varepsilon_n + \alpha\varepsilon_{n-1} + \cdots + \alpha^{n-1}\varepsilon_1 \sim N(0,(1-\alpha^{2n})\sigma^2/(1-\alpha^2)),$$

故

$$\begin{aligned} &P(x_n < a \mid x_0 = x) \\ &= P(\varepsilon_n + \alpha\varepsilon_{n-1} + \cdots + \alpha^{n-1}\varepsilon_1 < a - \alpha^n x_0) \\ &= \left[\sqrt{2\pi}\left(\frac{1-\alpha^{2n}}{1-\alpha^2}\sigma^2\right)^{\frac{1}{2}}\right]^{-1}\int_{-\infty}^{a}\exp\left[\frac{(z-\alpha^n x)}{2\left(\frac{1-\alpha^{2n}}{1-\alpha^2}\sigma^2\right)}\right]\mathrm{d}z. \end{aligned}$$

上式表明(4.2.10)式的 n 步转移概率可通过正态 $N(\alpha^n x,(1-\alpha^{2n})\sigma^2/(1-\alpha^2))$求出来.

当(4.2.10)式中 ε_t 的概率分布有密度 g_ε 时,由(4.2.10)式可得

$$\begin{aligned} P(x,(-\infty,a)) &= \int_{-\infty}^{a-\alpha x} g_\varepsilon(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{a} g_\varepsilon(z-\alpha x)\,\mathrm{d}z. \end{aligned}$$

上式表明在给定 $x_{t-1}=x$ 的条件下，x_t 的转移概率密度为将 ε_t 的概率密度作了 αx 的位移. 类似地，也可给出模型(4.2.10)的 n 步转移概率的表达式

$$\begin{aligned} & P(x_n < y \mid x_0 = x) \\ = & P(\varepsilon_n + \alpha\varepsilon_{n-1} + \cdots + \alpha^{n-1}\varepsilon_1 < y - \alpha^n x) \\ = & \int\cdots\int_{y_1+\alpha y_2+\cdots+\alpha^{n-1}y_n < y-\alpha^n x} g_\varepsilon(y_1)\cdots g_\varepsilon(y_n)\mathrm{d}y_1\cdots\mathrm{d}y_n. \end{aligned}$$

由§1.5的讨论知，模型(4.2.10)有平稳解

$$x_t = \sum_{j=0}^{\infty}\alpha^j\varepsilon_{t-j}.$$

取 $t=0$，有

$$x_0 = \sum_{j=0}^{\infty}\alpha^j\varepsilon_{-j}. \tag{4.2.11}$$

当 ε_t 的概率分布给定时，(4.2.11)式的概率分布是唯一确定的，且为模型(4.2.10)的不变概率分布. 为求模型(4.2.10)的不变概率分布，可利用随机变量的特征函数法，由 $\{\varepsilon_t\}$ 的独立及同分布性质和(4.2.11)式，可得

$$f_x(s) = \prod_{j=0}^{\infty} f_\varepsilon(\alpha^j s), \tag{4.2.12}$$

式中 $f_x(s)$ 和 $f_\varepsilon(s)$ 分别为 x_0 和 ε_t 的特征函数. 当 $\varepsilon_t \sim N(0,\sigma^2)$ 时，由(4.2.12)式不难求得

$$f_x(s) = e^{-\frac{1}{2}\frac{\sigma^2}{1-\alpha^2}s^2},$$

即 $x_0 \sim N(0,\sigma^2/(1-\alpha^2))$，但是对一般情况，利用(4.2.12)式给出 $f_x(s)$ 的明显表达式并不容易.

求模型(4.2.10)的不变概率分布还可直接使用(4.2.10)式. 记 F_x 为 x_t 的分布，F_ε 为 ε_t 的分布. 若 F 为模型(4.2.10)的不变概率分布，则由(4.2.10)式知

$$\begin{aligned} F(a) &= P(x_t < a) = P(\alpha x_{t-1} + \varepsilon_t < a) \\ &= \int_{-\infty}^{\infty} P(\alpha x_{t-1} + \varepsilon_t < a \mid x_{t-1} = y)\mathrm{d}F(y) \\ &= \int_{-\infty}^{\infty} P(\varepsilon_t < a - \alpha y)\mathrm{d}F(y) \\ &= \int_{-\infty}^{\infty} F_\varepsilon(a - \alpha y)\mathrm{d}F(y), \end{aligned}$$

即

$$F_x(a) = \int_{-\infty}^{\infty} F_\varepsilon(a - \alpha y)\mathrm{d}F_x(y). \tag{4.2.13}$$

此式表明了(4.2.10)式的不变概率分布应当满足(4.2.13)式的积分方程. 特别当 F_x 和 F_ε 分别有密度 g_x 和 g_ε 时,(4.2.13)式可写为

$$g_x(a) = \int_{-\infty}^{\infty} g_\varepsilon(a - \alpha y)g_x(y)\mathrm{d}y. \tag{4.2.14}$$

如果 F_ε 没有密度,例如 F_ε 为离散分布时,为简单起见,这里仅考虑 F_ε 服从两点分布的情况. 设

$$P(\varepsilon_t = -1) = P(\varepsilon_t = 1) = \frac{1}{2},$$

并注意到 ε_t 与 x_{t-1} 独立,由(4.2.13)式,有

$$\begin{aligned} F_x(a) &= P(\alpha x_{t-1} + \varepsilon_t < a) \\ &= \frac{1}{2}P(\alpha x_t < a + 1) + \frac{1}{2}P(\alpha x_t < a - 1). \end{aligned}$$

当 $0<\alpha<1$ 时,上式为

$$F_x(a) = \frac{1}{2}\{F_x((a+1)/\alpha) + F_x((a-1)/\alpha)\}.$$

显然,这一方程的解与 α 有密切的关系,但不易求解.

值得指出的是:尽管在一般情况下,给出模型(4.2.10)的不变概率分布的明显表达式不容易,但是只要 $\alpha\neq 0$, ε_t 的概率分布不退化,则模型(4.2.10)的不变概率分布就一定是连续的(见文献[119]). 注意:这里是指 x_0 的分布函数为连续的,并非 x_0 一定有密度函数. 它可以是 Lebesgue 测度奇异的. 作为练习,读者不妨考虑 ε_t 为二点分布的情况. 适当选取 α 值和二点分布的参数,便可发现(4.2.10)式有相对于 Lebesgue 测度奇异的连续分布.

由上述讨论可见,虽然模型(4.2.10)是齐时马氏链,但是不使用马氏链方法,也可以直接用模型本身对不变概率分布的存在性进行讨论,而且讨论的方法不难推广到 LAR(p)模型的情况. 总之,在研究 LAR 模型的概率结构时,使用马氏链的工具并不重要.

但是,对于 NLAR 模型(4.2.1),求出其转移概率(4.2.9)并不像模型(4.2.10)那样简单,而且也不能断言模型(4.2.1)的不变概率分布一定是连续函数. 例如对形如(3.2.1)的 NLAR 模型,就很难求出其一步转移概率的表达式. 好在当我们借助于一般状态空间马氏链的工具从理论上讨论 NLAR 模型(4.2.1)的几何遍历性时,并不需要其转移概率的明显

表达式. 论证了 NLAR 模型的遍历性，其有唯一平稳解则是必然的结论.

在这一小节的最后，我们讨论一元 p 阶 NLAR 模型与马氏链的联系. 考虑如下 NLAR 模型

$$\begin{cases} x_t = \varphi(x_{t-1},\cdots,x_{t-p},\varepsilon_t), & t \geqslant 1; \\ (x_0,x_{-1},\cdots,x_{-p+1})^\tau \in R^p. \end{cases} \tag{4.2.15}$$

其中$\{\varepsilon_t\}$为白噪声序列，满足

$$\begin{cases} E\varepsilon_t = 0, \ E\varepsilon_t^2 = \sigma^2 < \infty; \\ \varepsilon_t \text{ 与 } \{x_s, s<t\} \text{ 独立}. \end{cases} \tag{4.2.16}$$

令

$$\boldsymbol{x}_{t-1} = (x_{t-1},\cdots,x_{t-p})^\tau,$$

$$\boldsymbol{\varphi}(\boldsymbol{x}_{t-1},\varepsilon_t) = (\varphi(\boldsymbol{x}_{t-1},\varepsilon_t), x_{t-1},\cdots,x_{t-p+1})^\tau,$$

则模型(4.2.15)可写成

$$\begin{cases} \boldsymbol{x}_t = \boldsymbol{\varphi}(\boldsymbol{x}_{t-1},\varepsilon_t); \\ \boldsymbol{x}_0 \in R^p. \end{cases} \tag{4.2.17}$$

上式表明$\{\boldsymbol{x}_t\}$是一齐时马氏链，状态空间为$(R^p, \mathscr{B}_p)$. 需要指出：(4.2.15)式的$\{x_t\}$不是本节定义的马氏链. 但是由$\{x_t\}$的相邻的 p 个随机变量构成的向量随机序列 $\boldsymbol{x}_t = (x_t,\cdots,x_{t-p+1})^\tau$ 却是马氏链.

按照一维随机序列平稳性的定义 1.1.1，模型(4.2.15)的平稳解$\{x_t\}$的任何有限维分布是时间推移不变的. 特别是$(x_1,x_0,\cdots,x_{-p+2})^\tau$ 与 $(x_0,x_{-1},\cdots,x_{-p+1})^\tau$具有相同的概率分布. 由此可见，模型(4.2.15)的不变概率分布应定义为使$(x_0,x_{-1},\cdots,x_{-p+1})^\tau$ 与$(x_1,x_0,\cdots,x_{-p+2})^\tau$ 具有相同的概率分布，这正是模型(4.2.17)的不变概率分布. 由此可见，如果$\{\boldsymbol{x}_t\}$是模型(4.2.17)的平稳解，则由它确定的随机序列$\{x_t\}$便是(4.2.15)的平稳解. 反之，若$\{x_t, t\geqslant -p+1\}$是模型(4.2.15)的平稳解，则$\{\boldsymbol{x}_t, t\geqslant 0\}$就是模型(4.2.17)的平稳解. 在后文中，称模型(4.2.15)为几何遍历的，乃是指模型(4.2.17)为几何遍历的.

三、关于不可约、非周期和小集的引理

借助漂移准则判断由模型(4.2.1)所决定的马氏链的几何遍历性，是我们研究 NLAR 模型(4.2.1)的遍历性和平稳性的重要途径之一. 为了利用漂移准则，首先需要研究由模型(4.2.1)所决定的马氏链的不可约

性、非周期性，以及怎样的集合是小集的问题. 下面的两个引理对两类很重要的 NLAR 模型给出了其为不可约、非周期的条件，并回答了哪些集合是小集的问题.

引理 4. 2. 5　设时间序列$\{x_t\}$服从如下 NLAR 模型

$$\begin{cases} x_t = \varphi(x_{t-1}, \cdots, x_{t-p}) + \varepsilon_t; \\ (x_0, x_{-1}, \cdots, x_{-p+1})^\tau \in R^p, \end{cases} \tag{4.2.18}$$

其中$\{\varepsilon_t\}$为白噪声序列，φ是R^p到R^1的可测函数. 如果

(i) φ在R^p中的有界集上有界，即对每个正数$K<\infty$，有

$$\sup_{\|x\|\leqslant K} |\varphi(\boldsymbol{x})| < \infty. \tag{4.2.19}$$

(ii) 随机变量ε_t的概率分布为F_ε，其测度分解中含有绝对连续部分，其密度记为f_ε，它是处处为正的连续函数，而且ε_t与$\{x_s, s<t\}$相互独立，则模型(4. 2. 18)所决定的马氏链是齐时的、关于 Lebesgue 测度μ_p-不可约的、非周期的，而且R^p中任何有界μ_p-正测集都是小集.

证明　由对模型(4. 2. 15)的讨论知，模型(4. 2. 18)所决定的马氏链是齐时的.

以下仅对$p=2$来证明引理的其他结论，$p>2$的情况与之类似，除了繁琐外，别无本质差别.

先取$A=[a_1,b_1)\times[a_2,b_2)\in\mathscr{B}_2$，由(4. 2. 18)式有

$$x_1 = \varphi(x_0, x_{-1}) + \varepsilon_1,$$

$$x_2 = \varphi(\varphi(x_0, x_{-1}) + \varepsilon_1, x_0) + \varepsilon_2.$$

记$\boldsymbol{z}=(z_1,z_2)^\tau$，由条件(i)和(ii)、条件概率的性质及$\varepsilon_1$和$\varepsilon_2$的独立性，有

$$\begin{aligned} &P_2(z, A) \\ &= P(x_1 \in [a_1,b_1), x_2 \in [a_2,b_2) \mid x_0 = z_1, x_{-1} = z_2) \\ &= P(\varphi(z_1,z_2) + \varepsilon_1 \in [a_1,b_1), \varphi(\varphi(z_1,z_2) + \varepsilon_1, z_1) \\ &\quad + \varepsilon_2 \in [a_2,b_2)) \\ &= \int_{R^1} E\{I(\varphi(\varphi(z_1,z_2) + u_1, z_1) \\ &\quad + \varepsilon_2 \in [a_2,b_2)) \mid \varepsilon_1 = u_1\} EI(\varphi(z_1,z_2) \\ &\quad + \varepsilon_1 \in [a_1,b_1)) P_{\varepsilon_1}(\mathrm{d}u_1) \\ &\geqslant \int_{a_1-\varphi(z_1,z_2)}^{b_1-\varphi(z_1,z_2)} E\{I(\varphi(\varphi(z_1,z_2) + u_1, z_1) \\ &\quad + \varepsilon_2 \in [a_2,b_2)) \mid \varepsilon_1 = u_1\} \cdot f_\varepsilon(u_1)\mathrm{d}u_1 \end{aligned}$$

$$\geqslant \int_{a_1-\varphi(z_1,z_2)}^{b_1-\varphi(z_1,z_2)} \int_{a_2-\varphi(\varphi(z_1,z_2)+u_1,z_1)}^{b_2-\varphi(\varphi(z_1,z_2)+u_1,z_1)} f_\varepsilon(u_2)\mathrm{d}u_2 f_\varepsilon(u_1)\mathrm{d}u_1$$

$$= \int_{a_1}^{b_1}\int_{a_2}^{b_2} f_\varepsilon(y_1-\varphi(z_1,z_2))f_\varepsilon(y_2-\varphi(y_1,z_1))\mathrm{d}y_1\mathrm{d}y_2$$

$$= \iint_A f_\varepsilon(y_1-\varphi(z_1,z_2))f_\varepsilon(y_2-\varphi(y_1,z_1))\mathrm{d}y_1\mathrm{d}y_2. \quad (4.2.20)$$

故由引理的条件(i)、(ii)和(4.2.20)式知，对任何 $z\in R^2, A\in\mathscr{B}_2$，且 $\mu_2(A)>0$，都有

$$P_2(z,A)>0. \quad (4.2.21)$$

由定义 4.1.1 知，满足(4.2.18)式的马氏链是 μ_2 不可约的.

由(4.2.20)式的推导，还可证明：

$$P_3(z,A)>0,\ \forall\, z\in R^2,\ A\in\mathscr{B}_2, \text{且 } \mu_2(A)>0. \quad (4.2.22)$$

由(4.2.21)和(4.2.22)两式及引理 4.1.6 知，由(4.2.18)式决定的马氏链是非周期的.

最后，我们来证明 R^2 中的任何有界 μ_2-正测集都是小集. 设 $A\in\mathscr{B}_2$ 是有界集，且 $\mu_2(A)>0$，又设 $B\subset A$，且 $B\in\mathscr{B}_2, \mu_2(B)>0$，再记 $y=(y_1,y_2)^{\mathrm{T}}$，则由引理的条件(i)和(ii)知：

$$\inf_{y\in A, z\in A} f_\varepsilon(y_1-\varphi(z_1,z_2))f_\varepsilon(y_2-\varphi(y_1,z_1))>0. \quad (4.2.23)$$

将(4.2.23)代入(4.2.20)式，得

$$\inf_{z\in A} P_2(z,B)\geqslant \iint_B \inf_{z\in A, y\in B} f_\varepsilon(y_1,\varphi(z_1,z_2))f_\varepsilon(y_2,\varphi(y_1,z_1))\mathrm{d}y_1\mathrm{d}y_2>0.$$

再由 B 的任意性及引理 4.1.8，即知 A 是小集. ▌

在引理 4.2.5 中，虽然只对 $p=2$ 的情况给出了证明，而对于 $p>2$ 的情况，证明完全类似. 这里唯一应当注意的是：当 $p=2$ 时，

$$A=[a_1,b_1)\times[a_2,b_2),$$

当 ε_t 存在密度 f_ε 时，则有

$$\begin{aligned}
P(z,A) &= P(x_1\in A\,|\,x_0=z)\\
&= P\left(\begin{pmatrix}x_0\\x_1\end{pmatrix}\in A\,\middle|\,\begin{pmatrix}x_0\\x_{-1}\end{pmatrix}=\begin{pmatrix}z_1\\z_2\end{pmatrix}\right)\\
&= P(a_1\leqslant x_1<b_1,\ a_0\leqslant x_0<b_0\,|\,x_0=z_1,\ x_{-1}=z_2)\\
&= P(a_1\leqslant\varphi(x_0,x_{-1})+\varepsilon_1<b_1,\\
&\qquad a_0\leqslant x_0<b_1\,|\,x_0=z_1,\ x_{-1}=z_2)\\
&= P(a_1\leqslant\varphi(z_1,z_2)+\varepsilon_1<b_1,\ a_0\leqslant z_1<b_0)
\end{aligned}$$

$$= \int_{a_1}^{b_1} f_\varepsilon(y_1 - \varphi(z_1, z_2)) I(a_0 \leqslant z_1 < b_0) \mathrm{d}y_1$$

$$= \begin{cases} \int_{a_1 - \varphi(z_1,z_2)}^{b_1 - \varphi(z_1,z_2)} P_\varepsilon(\mathrm{d}u), & \text{当 } z_1 \in [a_0, b_0); \\ 0, & \text{当 } z_1 \overline{\in} [a_0, b_0). \end{cases}$$

所以在引理 4.2.5 的证明中，为了证明(4.1.13)，即不可约性及非周期性，我们考虑了 $P_2(z, A)$ 和 $P_3(z, A)$. 类似地，当 $p>2$ 时，应当考虑 $P_p(z,A)$ 和 $P_{p+1}(z,A)$.

另外，在引理 4.2.5 中，关于密度 f_ε 的假定可以放宽，例如只假定 f_ε 为分片连续，或者在任何有界集上的下界为正. 有关这些讨论，除了论证稍复杂以外，别无其他实质性困难.

引理 4.2.6　设时间序列 $\{x_t\}$ 服从如下 NLAR 模型：

$$\begin{cases} x_t = \varphi(x_{t-1}, \cdots, x_{t-p}) + \varepsilon_t S(x_{t-1}, \cdots, x_{t-p}); \\ (x_0, x_{-1}, \cdots, x_{-p+1})^\tau \in R^p, \end{cases} \tag{4.2.24}$$

其中 $\{\varepsilon_t\}$ 和 φ 满足引理 4.2.5 的条件. 如果 S 是 R^p 到 R^1 的正值连续函数，满足对每个实数 $K>0$，有

$$\sup_{\|x\| \leqslant K} S(x) < \infty, \qquad \inf_{\|x\| \leqslant K} S(x) > 0, \tag{4.2.25}$$

则由模型(4.2.24)决定的马氏链是齐时的、μ_p-不可约的、非周期的，且 R^p 中的任何有界 μ_p-正测集都是小集.

证明　显然，由(4.2.24)式决定的马氏链是齐时的. 类似于引理 4.2.5 的证明，在此仅考虑 $p=2$ 的情形.

首先，完全类似于(4.2.20)式的推导，可得到对任意的 $z \in R^2$ 及 $A \in \mathscr{B}_2$ 且 $\mu_2(A)>0$，有

$$P_2(z, A) \geqslant \iint_A f_\varepsilon\left(\frac{y_2 - \varphi(y_1, z_1)}{S(y_1, z_1)}\right) f_\varepsilon\left(\frac{y_1 - \varphi(z_1, z_2)}{S(z_1, z_2)}\right)$$

$$\times \frac{1}{S(y_1, z_1)} \cdot \frac{1}{S(z_1, z_2)} \mathrm{d}y_1 \mathrm{d}y_2.$$

注意到 φ 和 f_ε 满足引理 4.2.5 的条件(i)、(ii)，且 $S(y_1, z_1)$ 和 $S(z_1, z_2)$ 是恒正的，且满足(4.2.25)式，从而不影响引理 4.2.5 中论证的不可约性、非周期性以及关于小集的推理. 这样，仿照引理 4.2.5 的证明可证得本引理.　▌

应用引理 4.2.5 和引理 4.2.6，可证明许多 NLAR 模型所决定的马氏链是 μ-不可约的、非周期的，而且任何有界的 Lebesgue 正测集是小

集.这为我们运用几何遍历漂移准则作了很好的准备.尤其是关于小集的结果,它使得不太容易理解的小集概念在使用中变得很容易处理了.当然,以上两个引理,特别是引理 4.2.6 还有改进的余地.这有待于今后进一步研究.对于一般的 NLAR 模型(4.2.1),有如下引理,其证明见文献[134].

引理 4.2.7 设时间序列$\{\boldsymbol{x}_t\}$服从如下 NLAR 模型:

$$\begin{cases}\boldsymbol{x}_t = \boldsymbol{\varphi}(\boldsymbol{x}_{t-1}, \boldsymbol{\varepsilon}_t); \\ \boldsymbol{x}_0 \in R^m,\end{cases} \tag{4.2.26}$$

式中$\{\boldsymbol{\varepsilon}_t\}$是 m 维向量白噪声序列,$\boldsymbol{\varepsilon}_t$ 有下半连续的密度函数 f_ε,且 f_ε处处大于零,$\boldsymbol{\varphi}$ 是 $R^m \times R^m$ 到 R^m 的可测函数,满足下列条件:

(i) $\partial\boldsymbol{\varphi}(\boldsymbol{x},\boldsymbol{y})/\partial\boldsymbol{x}$ 和 $\partial\boldsymbol{\varphi}(\boldsymbol{x},\boldsymbol{y})/\partial\boldsymbol{y}$ 为 $R^m \times R^m$ 上的连续函数,且对任意给定的 $\boldsymbol{x}\in R^m$,$\boldsymbol{\varphi}(\boldsymbol{x},\cdot)$恰好把 R^m 映到 R^m;

(ii) $\left|\dfrac{\partial\boldsymbol{\varphi}(\boldsymbol{x},\boldsymbol{y})}{\partial\boldsymbol{y}}\right| \neq 0, \quad \forall\, \boldsymbol{x}、\boldsymbol{y}\in R^m.$

则由(4.2.26)式决定的马氏链是 μ_m-不可约的、非周期的,且 R^m 中的有界 μ_m-正测集都是小集.

四、NLAR 模型的几何遍历性

首先讨论可加噪声 NLAR 模型的几何遍历性,即考虑如下模型:

$$\begin{cases}x_t = \varphi(x_{t-1}, \cdots, x_{t-p}) + \varepsilon_t; \\ (x_0, x_{-1}, \cdots, x_{-p+1})^\tau \in R^p,\end{cases} \tag{4.2.27}$$

式中$\{\varepsilon_t\}$是白噪声序列,不一定要求 $E\varepsilon_t^2<\infty$,但满足

$$\begin{cases}E\varepsilon_t = 0,\ E|\varepsilon_t| < \infty; \\ \varepsilon_t \text{ 与 } \{x_s, s < t\} \text{ 独立.}\end{cases} \tag{4.2.28}$$

由§3.5 的讨论知,(4.2.27)可写成如下一阶 p 维 NLAR 模型:

$$\begin{cases}\boldsymbol{x}_t = \boldsymbol{\varphi}(\boldsymbol{x}_{t-1}) + \varepsilon_t\boldsymbol{u}; \\ \boldsymbol{x}_0 \in R^p,\end{cases} \tag{4.2.29}$$

式中 $\boldsymbol{x}_t = (x_t, \cdots, x_{t-p+1})^\tau$, $\boldsymbol{u} = (1, 0, \cdots, 0)^\tau$,且

$$\boldsymbol{\varphi}(\boldsymbol{x}_{t-1}) = (\varphi(\boldsymbol{x}_{t-1}), x_{t-1}, \cdots, x_{t-p+1})^\tau.$$

在(4.2.29)式中以 $\mathbf{0}$(零向量)代替 $\boldsymbol{u}$,便得到如下的非随机动态系统:

$$\begin{cases}\boldsymbol{y}_t = \boldsymbol{\varphi}(\boldsymbol{y}_{t-1}); \\ \boldsymbol{y}_0 \in R^p.\end{cases} \tag{4.2.30}$$

(4.2.30)式称为模型(4.2.29)的骨架模型，而(4.2.29)式犹如给骨架模型(4.2.30)添加了 $\varepsilon_t \boldsymbol{u}$ 的外衣. 以下定理与骨架模型(4.2.30)有密切联系.

定理 4.2.8　如果 NLAR 模型(4.2.29)满足下列两个条件：

(i) 存在某个 p 维向量范数 $\|\cdot\|_v$，常数 $0<\rho<1$ 和 $c\geqslant 0$，使得

$$\|\boldsymbol{\varphi}(\boldsymbol{x})\|_v \leqslant \rho \|\boldsymbol{x}\|_v + c,\quad \forall\, \boldsymbol{x} \in R^p; \tag{4.2.31}$$

(ii) ε_t 有处处为正的连续分布密度.

则模型(4.2.29)是几何遍历的.

证明　首先，由引理 4.2.5 知，在本定理的条件下，由模型(4.2.29)决定的马氏链 $\{\boldsymbol{x}_t\}$ 是 μ_p-不可约的、非周期的，且 R^p 中任何有界 μ_p-正测集都是小集. 这样，根据定理 4.1.12，为证明本定理的结论，只需证明：存在一个准则函数 g、一个有界集 C 和正数 $\lambda<1$，$c_1>0$，$c_2>0$，使得(4.1.25)和(4.1.26)两式成立. 为此，取 $g(\boldsymbol{x})=\|\boldsymbol{x}\|_v$. 由(4.2.28)、(4.2.29)和(4.2.31)诸式，有

$$\begin{aligned}
E\{g(\boldsymbol{x}_1)|\boldsymbol{x}_0=\boldsymbol{x}\} &= E\{g(\boldsymbol{\varphi}(\boldsymbol{x})+\varepsilon_1\boldsymbol{u})\}\\
&= E\|\boldsymbol{\varphi}(\boldsymbol{x})+\varepsilon_1\boldsymbol{u}\|_v\\
&\leqslant \|\boldsymbol{\varphi}(\boldsymbol{x})\|_v + E\|\varepsilon_1\boldsymbol{u}\|_v\\
&\leqslant \rho\|\boldsymbol{x}\|_v + c + c'\\
&\leqslant \lambda\|\boldsymbol{x}\|_v - [(\lambda-\rho)\|\boldsymbol{x}\|_v - c - c'],
\end{aligned} \tag{4.2.32}$$

式中 $0<\rho<\lambda<1$，$c'=(E|\varepsilon_1|)\|\boldsymbol{u}\|_v$. 取有界集

$$C=\{\boldsymbol{x}:\ \|\boldsymbol{x}\|\leqslant K\},\ K>(c+c')/(\lambda-\rho),$$

$$c_1=(\lambda-\rho)K-c-c',\ c_2=\rho K+c+c',$$

则由(4.2.32)式得到

$$\begin{aligned}
E\{g(\boldsymbol{x}_1)|\boldsymbol{x}_0=\boldsymbol{x}\} &\leqslant \lambda\|\boldsymbol{x}\|_v - c_1\\
&= \lambda g(\boldsymbol{x}) - c_1,\quad \forall\, \boldsymbol{x}\,\overline{\in}\, C;\\
E\{g(\boldsymbol{x}_1)|\boldsymbol{x}_0=\boldsymbol{x}\} &\leqslant c_2,\qquad \forall\, x\in C.
\end{aligned}$$

至此定理得证.　▌

推论 4.2.9　若 NLAR 模型(4.2.27)满足下列两条件：

(i) 存在常数向量 $\boldsymbol{\alpha}=(\alpha_1,\cdots,\alpha_p)^\tau$，满足

$$1-\alpha_1 z-\cdots-\alpha_p z^p \neq 0,\quad |z|\leqslant 1, \tag{4.2.33}$$

使得

$$\lim_{\|\boldsymbol{x}\|\to\infty}\frac{|\varphi(\boldsymbol{x})-\boldsymbol{\alpha}^\tau\boldsymbol{x}|}{\|\boldsymbol{x}\|}=0, \tag{4.2.34}$$

且对每一个 $K>0$，

$$\sup_{\|x\|\leqslant K}|\varphi(x)-\alpha^{\tau}x|<\infty; \tag{4.2.35}$$

(ii) ε_t 有处处为正的连续分布密度，

则 NLAR 模型(4.2.27)是几何遍历的.

证明 由定理 4.2.8，只需证明在推论的条件下，模型(4.2.27)的一阶 p 维变形(4.2.29)式中的 $\boldsymbol{\varphi}$ 满足条件(4.2.31)式即可. 为此，令

$$A=\begin{pmatrix}\alpha_1 & \cdots & \alpha_{p-1} & \alpha_p\\ 1 & \cdots & 0 & 0\\ \vdots & \ddots & \vdots & \vdots\\ 0 & \cdots & 1 & 0\end{pmatrix},$$

$$\begin{aligned}\boldsymbol{H}(\boldsymbol{x})&=\boldsymbol{\varphi}(\boldsymbol{x})-A(\boldsymbol{x})\\&=(\varphi(\boldsymbol{x})-\boldsymbol{\alpha}^{\tau}\boldsymbol{x},0,\cdots,0)^{\tau}.\end{aligned}$$

于是

$$\boldsymbol{\varphi}(\boldsymbol{x})=\boldsymbol{H}(\boldsymbol{x})+A\boldsymbol{x}. \tag{4.2.36}$$

由(4.2.33)式和推论 2.3.10 知，$\rho(A)<1$，从而由定理 2.3.9 知，存在一个 p 维向量范数 $\|\cdot\|_v$，使得变换 $\boldsymbol{T}(\boldsymbol{x})=A\boldsymbol{x}$ 为依范数 $\|\cdot\|_v$ 压缩的，即存在正数 $\rho<1$，使得

$$\|A\boldsymbol{x}\|_v\leqslant\rho\|\boldsymbol{x}\|_v,\quad \forall\,\boldsymbol{x}\in R^p. \tag{4.2.37}$$

而由 p 维向量范数的等价性，有

$$\frac{\|\boldsymbol{H}(\boldsymbol{x})\|_v}{\|\boldsymbol{x}\|_v}\leqslant\frac{c\|\boldsymbol{H}(\boldsymbol{x})\|}{\|\boldsymbol{x}\|}=\frac{c|\varphi(\boldsymbol{x})-\boldsymbol{\alpha}^{\tau}\boldsymbol{x}|}{\|\boldsymbol{x}\|},$$

从而由(4.2.34)式，有

$$\lim_{\|\boldsymbol{x}\|_v\to\infty}\frac{\|\boldsymbol{H}(\boldsymbol{x})\|_v}{\|\boldsymbol{x}\|_v}=0.$$

由此式可知，必存在一个正数 $K_0>0$，使得

$$\frac{\|\boldsymbol{H}(\boldsymbol{x})\|_v}{\|\boldsymbol{x}\|_v}<\frac{1}{2}(1-\rho),\quad 当\ \|\boldsymbol{x}\|_v\geqslant K_0. \tag{4.2.38}$$

由(4.2.35)式和向量范数的等价性，有

$$\begin{aligned}\|\boldsymbol{\varphi}(\boldsymbol{x})\|_v&\leqslant\|A\boldsymbol{x}\|_v+\|\boldsymbol{H}(\boldsymbol{x})\|_v\\&\leqslant\rho\|\boldsymbol{x}\|_v+c_1\|\boldsymbol{H}(\boldsymbol{x})\|\\&=\rho\|\boldsymbol{x}\|_v+c_1|\varphi(\boldsymbol{x})-\boldsymbol{\alpha}^{\tau}\boldsymbol{x}|\\&\leqslant\rho K_0+c_2\\&\leqslant c<\infty,\quad 当\ \|\boldsymbol{x}\|_v\leqslant K_0,\end{aligned}$$

式中 $c>0$ 为常数，且在不同地方出现，其含义也可不同. 由(4.2.37)和(4.2.38)两式有

$$\begin{aligned}\|\boldsymbol{\varphi}(\boldsymbol{x})\|_v &\leqslant \|A\boldsymbol{x}\|_v + \|\boldsymbol{H}(\boldsymbol{x})\|_v \\ &\leqslant \rho\|\boldsymbol{x}\|_v + \frac{1}{2}(1-\rho)\|\boldsymbol{x}\|_v \\ &= \frac{1}{2}(1+\rho)\|\boldsymbol{x}\|_v, \quad \text{当 } \|\boldsymbol{x}\|_v \geqslant K_0.\end{aligned}$$

综合上述两式得

$$\|\boldsymbol{\varphi}(\boldsymbol{x})\|_v \leqslant \lambda\|\boldsymbol{x}\|_v + c, \quad \forall\, \boldsymbol{x} \in R^p.$$

此即(4.2.31)式. 从而推论得证. ▌

推论 4.2.10　设时间序列$\{x_t\}$服从 LAR(p)模型，即

$$x_t = \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + \varepsilon_t, \tag{4.2.39}$$

式中系数向量$\boldsymbol{\alpha}=(\alpha_1,\cdots,\alpha_p)^{\tau}$满足(4.2.33)式，$\{\varepsilon_t\}$为白噪声序列，满足(4.2.28)式及定理 4.2.8 的条件(ii)，则模型(4.2.39)是几何遍历的.

证明　推论的证明已蕴含于推论 4.2.9 的证明中，这里从略. ▌

对上述定理及推论，我们给出如下注解和例子：

注 4.2.11　当模型(4.2.39)为平稳 LAR 模型时，相应的一阶 p 维 LAR 模型为

$$\boldsymbol{x}_t = A\boldsymbol{x}_{t-1} + \varepsilon_t\boldsymbol{u}, \tag{4.2.40}$$

式中 A 如推论 4.2.9 的证明中所定义，其谱半径 $\rho(A)<1$. 于是，由(4.1.3)式可得

$$\begin{aligned}\boldsymbol{x}_t &= \varepsilon_t\boldsymbol{u} + A\varepsilon_{t-1}\boldsymbol{u} + \cdots + A^{t-1}\varepsilon_1\boldsymbol{u} + A^t\boldsymbol{x}_0 \\ &= \sum_{k=0}^{\infty} A^k\varepsilon_{t-k}\boldsymbol{u} + A^t\boldsymbol{x}_0 - A^t\sum_{j=0}^{\infty} A^j\varepsilon_{-j}\boldsymbol{u} \\ &= \boldsymbol{w}_t + A^t\boldsymbol{x}_0 - A^t\boldsymbol{w}_0,\end{aligned} \tag{4.2.41}$$

式中

$$\boldsymbol{w}_t = \sum_{j=0}^{\infty} A^j\varepsilon_{t-j}\boldsymbol{u}$$

是一平稳序列，$\boldsymbol{w}_t$ 的分布为模型(4.2.40)的不变分布. 不难看出，对 ε_t 的分布不需要定理 4.2.8 中那样强的限制，$\{\boldsymbol{x}_t\}$仍能有几何遍历性. 由于本书以介绍非线性时间序列模型为主，不讨论保证(4.2.40)式具有几何遍历性的更弱的条件，这里仅举一例，说明几何遍历性条件在适当的范数下可以减弱.

例 4.2.12 考虑如下 LAR(1)模型

$$x_t = \frac{1}{2}x_{t-1} + \varepsilon_t, \tag{4.2.42}$$

其中$\{\varepsilon_t\}$为白噪声序列，且

$$P\left(\varepsilon_t = \frac{1}{4}\right) = P\left(\varepsilon_t = -\frac{1}{4}\right) = \frac{1}{2}.$$

令 $y_t = \frac{1}{2} + x_t$，$\eta_t = \frac{1}{4} + \varepsilon_t$，则(4.2.42)式可写成

$$y_t = \frac{1}{2}y_{t-1} + \eta_t, \tag{4.2.43}$$

其中$\{\eta_t\}$为独立同分布序列，且 $P(\eta_t = 0) = P\left(\eta_t = \frac{1}{2}\right) = \frac{1}{2}$. 由(4.2.41)和(4.2.43)两式有

$$y_t = v_t + \left(\frac{1}{2}\right)^t y_0 - \left(\frac{1}{2}\right)^t v_0, \tag{4.2.44}$$

式中

$$v_t = \sum_{j=0}^{\infty}\left(\frac{1}{2}\right)^j \eta_{t-j}.$$

易见 $0 \leqslant v_t \leqslant 1$，$\{v_t\}$为平稳序列. 下面求 v_t 的分布：

$$P\left(v_t < \frac{1}{2}\right) = P(\varepsilon_t = 0) = \frac{1}{2},$$

$$\begin{aligned}
P\left(v_t < \frac{1}{2} + \frac{1}{4}\right) &= P\left(v_t < \frac{1}{2}\right) + P\left(\frac{1}{2} \leqslant v_t < \frac{1}{2} + \frac{1}{4}\right) \\
&= P(\varepsilon_t = 0) + P(\varepsilon_t = 1, \varepsilon_{t-1} = 0) \\
&= \frac{1}{2} + \frac{1}{4} = \frac{3}{4},
\end{aligned}$$

对于任意的有穷位二进制小数

$$a = \left(\frac{1}{2}\right)^{k_1} + \left(\frac{1}{2}\right)^{k_2} + \cdots + \left(\frac{1}{2}\right)^{k_s},$$

其中 $k_1 < k_2 < \cdots < k_s$ 为某些正整数，类似于上述推理，可得

$$\begin{aligned}
P(v_t < a) = &P\left(0 < v_t < \left(\frac{1}{2}\right)^{k_1}\right) \\
&+ P\left(\left(\frac{1}{2}\right)^{k_1} \leqslant v_t < \sum_{i=1}^{s}\left(\frac{1}{2}\right)^{k_i}\right) \\
= &P(\varepsilon_t = 0, \varepsilon_{t-1} = 0, \cdots, \varepsilon_{t-k_1} = 0) \\
&+ P\left(\left(\frac{1}{2}\right)^{k_1} \leqslant v_t < \sum_{i=1}^{s}\left(\frac{1}{2}\right)^{k_i}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \left(\frac{1}{2}\right)^{k_1} + P\left(\left(\frac{1}{2}\right)^{k_1} \leqslant v_t < \left(\frac{1}{2}\right)^{k_1} + \left(\frac{1}{2}\right)^{k_2}\right) \\
&\quad + P\left(\left(\frac{1}{2}\right)^{k_1} + \left(\frac{1}{2}\right)^{k_2} \leqslant v_t < \sum_{i=1}^{s}\left(\frac{1}{2}\right)^{k_i}\right) = \cdots \\
&= \left(\frac{1}{2}\right)^{k_1} + \left(\frac{1}{2}\right)^{k_2} + \cdots + \left(\frac{1}{2}\right)^{k_s} \\
&= a.
\end{aligned}
$$

注意到有穷位二进制小数在[0,1]中为稠密的，且 v_t 的分布函数为连续的，从而由上式得

$$P(x_t < a) = a, \quad 0 \leqslant a \leqslant 1.$$

上式意味着 v_t 的分布为[0,1]上的均匀分布. 于是

$$
\begin{aligned}
P(y_t < x) &= P\left(v_t + \left(\frac{1}{2}\right)^t y_0 - \left(\frac{1}{2}\right)^t v_0 < x\right) \\
&= P\left(v_t < x + \left(\frac{1}{2}\right)^t v_0 - \left(\frac{1}{2}\right)^t y_0\right) \\
&\leqslant P\left(v_t < x + \left(\frac{1}{2}\right)^t (|y_0| + 1)\right).
\end{aligned}
$$

另一方面，

$$
\begin{aligned}
P(y_t < x) &= P\left(v_t + \left(\frac{1}{2}\right)^t y_0 - \left(\frac{1}{2}\right)^t v_0 < x\right) \\
&\geqslant P\left(v_t > x - \left(\frac{1}{2}\right)^t (|y_0| + 1)\right).
\end{aligned}
$$

以 F_U 表示[0,1]上均匀分布的分布函数，以 F_t 表示 y_t 的分布函数. 由上述两式不难算得

$$\sup_x |F_t(x) - F_U(x)| \leqslant \left(\frac{1}{2}\right)^t (3|y_0| + 3). \tag{4.2.45}$$

若用以下的范数代替(4.1.22)式中的全变差范数，即

$$\| F - G \|_\infty = \sup_x |F(x) - G(x)|, \tag{4.2.46}$$

则(4.2.45)表明了 F_t 在(4.2.46)式定义的范数下以指数速度收敛于 F_U. 如果用下式代替定义 4.2.4 中的(4.2.8)式，即

$$\lim_{t\to\infty} \rho^{-t} \| F_t - F_U \|_\infty = 0,$$

则在此意义下，模型(4.2.42)是几何遍历的. 但此时却不要求 ε_t 有处处为正的密度函数.

注 4.2.13 如果定理 4.2.8 中的条件(4.2.31)式中的 $c=0$，即

$$\|\boldsymbol{\varphi}(\boldsymbol{y})\|_v \leqslant \rho \|\boldsymbol{\varphi}(\boldsymbol{y})\|_v, \quad \forall\, \boldsymbol{y} \in R^p, \tag{4.2.47}$$

式中正数 $\rho<1$，则依定理 4.2.8，NLAR 模型(4.2.27)是几何遍历的. 实际上，$\boldsymbol{\varphi}$ 满足(4.2.47)式的压缩条件时，相应于(4.2.29)式的骨架模型(4.2.30)，其吸引集是只包含零点的单点集，而且零点是稳定的不动点. 如第 2 章所述，这样的动态系统的结构太简单了，与稳定的线性系统太相近了.

定理 4.2.14 如果 NLAR 模型(4.2.27)中的 $\{\varepsilon_t\}$ 满足定理 4.2.8 的条件，且存在常数 $0<\lambda<1$ 和 $c\geqslant 0$，使得

$$|\varphi(x_1,\cdots,x_p)| \leqslant \lambda \max\{|x_1|,\cdots,|x_p|\} + c, \tag{4.2.48}$$

则模型(4.2.27)是几何遍历的.

证明 首先由引理4.2.5知：在条件(4.2.48)下，由模型(4.2.29)决定的马氏链 $\{\boldsymbol{x}_t\}$ 是 μ_p-不可约的、非周期的，而且任何有界 μ_p-正测集都是小集. 取准则函数

$$g(\boldsymbol{x}) = \max\{|x_1|,\cdots,|x_p|\} = \|\boldsymbol{x}\|_\infty, \quad \forall\, x \in R^p,$$

由(4.2.27)式和(4.2.48)式有

$$\begin{aligned}
|x_1| &\leqslant |\varepsilon_1| + c + \lambda \max\{|x_0|,\cdots,|x_{-p+1}|\} \\
&= |\varepsilon_1| + c + \lambda \|\boldsymbol{x}_0\|_\infty, \\
|x_2| &\leqslant |\varepsilon_2| + c + \lambda \max\{|\varepsilon_1| + c \\
&\qquad + \lambda \max\{|x_0|,\cdots,|x_{-p+1}|\}, |x_0|,\cdots,|x_{-p+2}|\} \\
&\leqslant |\varepsilon_2| + c + \lambda(|\varepsilon_1| + c) \\
&\qquad + \lambda \max\{|x_0|,\cdots,|x_{-p+2}|, \lambda|x_{-p+1}|\} \\
&\leqslant |\varepsilon_2| + c + \lambda(|\varepsilon_1| + c) + \lambda\|\boldsymbol{x}_0\|_\infty.
\end{aligned}$$

依次类推可得

$$\begin{aligned}
|x_k| &\leqslant (|\varepsilon_k| + c) + \lambda(|\varepsilon_{k-1}| + c) + \cdots \\
&\qquad + \lambda^{k-1}(|\varepsilon_1| + c) + \lambda\|\boldsymbol{x}_0\|_\infty, \\
&\qquad (k = 1,2,\cdots,p).
\end{aligned} \tag{4.2.49}$$

于是由(4.2.49)式，有

$$\begin{aligned}
E\{g(\boldsymbol{x}_p) | \boldsymbol{x}_0 = \boldsymbol{x}\} &= E\{\max\{|x_p|,\cdots,|x_1|\} | \boldsymbol{x}_0 = \boldsymbol{x}\} \\
&\leqslant p(E|\varepsilon_1| + c) + \lambda\|\boldsymbol{x}_0\|_\infty.
\end{aligned} \tag{4.2.50}$$

对 $\{\boldsymbol{x}_t\}$ 的子序列 $\{\boldsymbol{x}_{tp}\}$，令 $\boldsymbol{y}_t = \boldsymbol{x}_{tp}$，则由(4.2.50)式有

$$\begin{aligned}
E\{g(\boldsymbol{y}_1) | \boldsymbol{y}_0 = \boldsymbol{x}\} &\leqslant \lambda\|\boldsymbol{x}\|_\infty + c' \\
&\leqslant \lambda'\|\boldsymbol{x}\|_\infty - [(\lambda' - \lambda)\|\boldsymbol{x}\|_\infty - c'],
\end{aligned} \tag{4.2.51}$$

其中$0<\lambda<\lambda'<1$，$c'=p(E|\varepsilon_1|+c)$. 利用$\lambda'-\lambda>0$，由(4.2.51)式和g的定义，存在正整数K_1，使当$\|\boldsymbol{x}\|_\infty\geqslant K_1$时，$(\lambda'-\lambda)g(\boldsymbol{x})-c'>0$. 取$K>K_1$，并令$A=\{\boldsymbol{x}:\|\boldsymbol{x}\|_\infty\leqslant K\}$，则由(4.2.51)式知：存在常数$c_1>0$和$c_2>0$，使得

$$E\{g(\boldsymbol{y}_1)|\boldsymbol{y}_0=x\}\leqslant\lambda' g(\boldsymbol{x})-c_1,\quad\forall\, x\overline{\in} A;$$

$$E\{g(\boldsymbol{y}_1)|\boldsymbol{y}_0=x\}\leqslant c_2,\quad\forall\, x\in A.$$

故由定理4.1.12知，$\{\boldsymbol{x}_{tp}\}$是几何遍历的. 再由定理4.1.13知，$\{\boldsymbol{x}_t\}$是几何遍历的，从而模型(4.2.27)是几何遍历的. 于是定理得证. ∎

设时间序列$\{x_t\}$服从 NLAR 模型(4.2.27). 对变换φ引进如下记号：

$$\begin{aligned}y_1&=\varphi(y_0,y_{-1},\cdots,y_{-p+1})\triangleq\varphi_1(y_0,y_{-1},\cdots,y_{-p+1}),\\ y_2&=\varphi(y_1,y_0,\cdots,y_{-p+2})\\ &=\varphi(\varphi(y_0,\cdots,y_{-p+1}),y_0,\cdots,y_{-p+2})\\ &\triangleq\varphi_2(y_0,y_{-1},\cdots,y_{-p+1}),\end{aligned}$$

一般地有

$$\begin{aligned}y_k&=\varphi(y_{k-1},\cdots,y_{k-p})=\varphi_2(y_{k-2},\cdots,y_{k-p-1})\\ &=\cdots=\varphi_k(y_0,y_{-1},\cdots,y_{-p+1}),\quad k\geqslant 1.\end{aligned}\tag{4.2.52}$$

定理 4.2.15　如果 NLAR 模型(4.2.27)中的$\{\varepsilon_t\}$满足定理4.2.8的条件，且变换φ满足下列条件组(i)、(ii)或(i)、(ii)′：

(i) 存在正数c_k，使得对任何$y_0',y_0,y_{-1},\cdots,y_{-p+1}$和$k\geqslant1$，都有

$$|\varphi_k(y_0,y_{-1},\cdots,y_{-p+1})-\varphi_k(y_0',y_{-1},\cdots,y_{-p+1})|\leqslant c_k|y_0-y_0'|;\tag{4.2.53}$$

(ii) 存在正常数$\rho<1$和M及c，使得对任何$y_0,y_{-1},\cdots,y_{-p+1}$和$k\geqslant 1$，有

$$|\varphi_k(y_0,y_{-1},\cdots,y_{-p+1})|\leqslant M\rho^k(|y_0|+\cdots+|y_{-p+1}|)+c;\tag{4.2.54}$$

(ii)′ 存在正常数$\rho<1$和c，使得对任何$y_0,y_{-1},\cdots,y_{-p+1}$和某个$m\geqslant1$，有

$$\sum_{k=1}^{p}|\varphi_{m+k}(y_0,\cdots,y_{-p+1})|\leqslant\lambda\left(\sum_{j=1}^{p}|y_{1-j}|\right)+c,\tag{4.2.55}$$

则模型(4.2.27)是几何遍历的.

证明　我们仅在条件(i)、(ii)下给出定理的证明. 在条件(i)、(ii)′下

的证明完全类似.

由(4.2.52)和(4.2.54)两式知,对任意的$(y_0,\cdots,y_{-p+1})\in R^p$,有

$$|\varphi(y_0,\cdots,y_{-p+1})|\leqslant M\rho(|y_0|+\cdots+|y_{-p+1}|)+c.$$

从而根据引理4.2.5知,由(4.2.29)式决定的马氏链$\{\boldsymbol{x}_t\}$是μ_p-不可约的、非周期的,而且任何有界的μ_p-正测度集都是小集,这里$\boldsymbol{x}_t=(x_t,x_{t-1},\cdots,x_{t-p+1})^\tau$.

由(4.2.53)和(4.2.54)两式,经反复迭代,可得

$$\begin{aligned}|x_t|&\leqslant|\varepsilon_t|+|\varphi(x_{t-1},\cdots,x_{t-p})|\\&=|\varepsilon_t|+|\varphi(\varepsilon_{t-1}+\varphi(x_{t-2},\cdots,x_{t-p-1}),x_{t-2},\cdots,x_{t-p})|\\&\leqslant|\varepsilon_t|+|\varphi(\varepsilon_{t-1}+\varphi(x_{t-2},\cdots,x_{t-p-1}),x_{t-2},\cdots,x_{t-p})\\&\quad-\varphi(\varphi(x_{t-2},\cdots,x_{t-p-1}),x_{t-2},\cdots,x_{t-p})|\\&\quad+|\varphi(\varphi(x_{t-2},\cdots,x_{t-p-1}),x_{t-2},\cdots,x_{t-p})|\\&\leqslant|\varepsilon_t|+c_1|\varepsilon_{t-1}|+|\varphi_2(x_{t-2},\cdots,x_{t-p-1})|\\&\leqslant\cdots\\&\leqslant|\varepsilon_t|+c_1|\varepsilon_{t-1}|+\cdots+c_{t-1}|\varepsilon_1|\\&\quad+|\varphi_t(x_0,\cdots,x_{-p+1})|\\&\leqslant|\varepsilon_t|+c_1|\varepsilon_{t-1}|+\cdots+c_{t-1}|\varepsilon_1|\\&\quad+M\rho^t(|x_0|+\cdots+|x_{-p+1}|)+c.\end{aligned}\tag{4.2.56}$$

记向量$\boldsymbol{v}=(v_1,\cdots,v_p)^\tau$,取准则函数

$$g(\boldsymbol{v})=|v_1|+\cdots+|v_p|\triangleq\|\boldsymbol{v}\|_1,\quad\forall\,\boldsymbol{v}\in R^p.$$

再选取足够大的正整数N,使得

$$pM\rho^{N-p}=\lambda<1.$$

于是,由(4.2.56)式,并注意到$E|\varepsilon_1|<\infty$,便得到

$$Eg(\boldsymbol{x}_N)\leqslant\lambda g(\boldsymbol{x}_0)+c',$$

式中$\{\boldsymbol{x}_t\}$满足模型(4.2.27)相应的一阶p维模型(4.2.29),且

$$c'=pc+\left(1+\sum_{i=1}^{N}c_i\right)E|\varepsilon_1|.$$

令$\boldsymbol{y}_t=\boldsymbol{x}_{Nt}$,则仿照定理4.2.14的后半部分证明,可证$\{\boldsymbol{x}_{Nt}\}$是几何遍历的,从而再由定理4.1.13知,模型(4.2.27)是几何遍历的. ▌

注 4.2.16　如果NLAR模型(4.2.27)中的变换φ满足

$$|\varphi(\boldsymbol{x})-\varphi(\boldsymbol{y})|\leqslant K\|\boldsymbol{x}-\boldsymbol{y}\|,\quad\forall\,\boldsymbol{x},\boldsymbol{y}\in R^p,\tag{4.2.57}$$

式中K为某个适当的正常数,当$p=1$时,φ满足定理4.2.15中的条

件(i).

推论 4.2.17　设 NLAR 模型(4.2.27)中的$\{\varepsilon_t\}$满足定理4.2.8的条件.如果存在R^p到R^1的可测函数ψ,以ψ代替定理4.2.15中的φ时,ψ满足定理4.2.15中的条件(i)、(ii)或(i)、(ii)$'$,且

$$|\varphi(\boldsymbol{x})-\psi(\boldsymbol{x})|\leqslant c,\quad \forall\, \boldsymbol{x}\in R^p, \tag{4.2.58}$$

其中c为非负常数,则模型(4.2.27)是几何遍历的.

证明　我们仅就ψ满足(4.2.53)和(4.2.54)给出证明,而ψ满足(4.2.53)和(4.2.55)时的证明是完全类似的.

由(4.2.54)和(4.2.58)两式,有

$$\begin{aligned}|\varphi(\boldsymbol{y})|&\leqslant c+|\psi(\boldsymbol{y})|\\&\leqslant c+M\rho^k(|y_1|+\cdots+|y_p|)+c',\\&\quad\forall\,\boldsymbol{y}=(y_1,\cdots,y_p)^\tau\in R^p.\end{aligned}$$

从而由引理4.2.5知,由模型(4.2.29)决定的马氏链$\{\boldsymbol{x}_t\}$是μ_p-不可约的、非周期的,而且任何有界μ_p-正测集是小集.

由(4.2.27)和(4.2.58)两式,有

$$|x_t-\psi(x_{t-1},\cdots,x_{t-p})|\leqslant|\varepsilon_t|+c. \tag{4.2.59}$$

对ψ沿用(4.2.52)式的记号,由(4.2.56)、(4.2.58)和(4.2.59)诸式,并注意用ψ_k代替(4.2.54)中的φ_k,又有

$$\begin{aligned}|x_t|&\leqslant|\varepsilon_t|+|\psi(x_{t-1},\cdots,x_{t-p})|+c\\&\leqslant|\varepsilon_t|+c\\&\quad+|\psi(\varepsilon_{t-1}+\varphi(x_{t-2},\cdots,x_{t-p-1}),x_{t-2},\cdots,x_{t-p})\\&\quad-\psi(\psi(x_{t-2},\cdots,x_{t-p-1}),x_{t-2},\cdots,x_{t-p})|\\&\quad+|\psi(\psi(x_{t-2},\cdots,x_{t-p-1}),x_{t-2},\cdots,x_{t-p})|\\&\leqslant|\varepsilon_t|+c+c_1(|\varepsilon_{t-1}|+c)+|\psi_2(x_{t-2},\cdots,x_{t-p-1})|\\&\leqslant\cdots\\&\leqslant(|\varepsilon_t|+c)+c_1(|\varepsilon_{t-1}|+c)+\cdots+c_{t-1}(|\varepsilon_1|+c)\\&\quad+|\psi_t(x_0,x_{-1},\cdots,x_{-p+1})|\\&\leqslant(|\varepsilon_t|+c)+\sum_{j=1}^{t-1}c_j(|\varepsilon_{t-j}|+c)\\&\quad+M\rho^t\Big(\sum_{j=1}^{p}|x_{1-j}|\Big)+c.\end{aligned}$$

以下仿照定理4.2.15证明的后半部分,即可证明本推论的结论.　▌

现在我们来讨论随机条件方差 NLAR 模型的几何遍历性. 设时间序列$\{x_t\}$服从函数型随机条件方差 NLAR 模型(3.3.20),即

$$x_t = \varphi(x_{t-1},\cdots,x_{t-p}) + \varepsilon_t S(x_{t-1},\cdots,x_{t-p}), \qquad (4.2.60)$$

其中$\{\varepsilon_t\}$为白噪声序列,满足(3.3.19)式,φ 和 S 都是 R^p 到 R^1的可测函数,且 S 是处处为正的. 模型(4.2.60)的一阶 p 维 NLAR 形式为

$$\begin{cases} \boldsymbol{x}_t = \boldsymbol{\varphi}(\boldsymbol{x}_{t-1}) + \varepsilon_t S(\boldsymbol{x}_{t-1})\boldsymbol{u}; \\ \boldsymbol{x}_0 \in R^p, \end{cases} \qquad (4.2.61)$$

式中 $\boldsymbol{x}_t$、$\boldsymbol{u}$ 和 $\boldsymbol{\varphi}$ 的定义与(4.2.29)式的相同.

定理 4.2.18 如果 NLAR 模型(4.2.60)满足下列条件:

(i) $\{\varepsilon_t\}$为白噪声序列,满足(3.3.19)式,且 ε_t 有处处为正的连续密度;

(ii) $\boldsymbol{\varphi}$ 满足定理4.2.8的条件(i);

(iii) $S(\boldsymbol{x})$是正值连续函数,满足对每个实数 $K>0$,

$$\sup_{\|\boldsymbol{x}\|\leqslant K} S(\boldsymbol{x}) < \infty, \quad \inf_{\|\boldsymbol{x}\|\leqslant K} S(\boldsymbol{x}) > 0, \qquad (4.2.62)$$

且

$$\lim_{\|\boldsymbol{x}\|\to\infty} \frac{S(\boldsymbol{x})}{\|\boldsymbol{x}\|} = 0, \qquad (4.2.63)$$

则模型(4.2.60)是几何遍历的.

证明 根据引理4.2.6知,由(4.2.60)决定的马氏链是 μ_p-不可约的、非周期的,而且任何有界 μ_p-正测集是小集. 根据条件(ii)取准则函数 $g(\boldsymbol{x}) = \|\boldsymbol{x}\|_v$,这里 $\boldsymbol{x}\in R^p$. 则由条件(i)和(ii)及 p 维向量范数的等价性,存在常数 $c_1>0$,使得

$$\begin{aligned} E\{g(\boldsymbol{x}_1)\mid \boldsymbol{x}_0 = \boldsymbol{x}\} &\leqslant E\|\boldsymbol{\varphi}(\boldsymbol{x}) + \varepsilon_1 S(\boldsymbol{x})\boldsymbol{u}\|_v \\ &\leqslant \|\boldsymbol{\varphi}(\boldsymbol{x})\|_v + \|S(\boldsymbol{x})\boldsymbol{u}\|_v \cdot E|\varepsilon_1| \\ &\leqslant \rho\|\boldsymbol{x}\|_v + (c_1 S(\boldsymbol{x})/\|\boldsymbol{x}\|_v)\cdot\|\boldsymbol{x}\|_v + c. \end{aligned} \qquad (4.2.64)$$

又由条件(iii)知,对于$\frac{1}{2}(\lambda-\rho)$,其中 $\rho<\lambda<1$,存在 $K_0>0$,使当$\|\boldsymbol{x}\|_v\geqslant K_0$时,有

$$\frac{c_1 S(\boldsymbol{x})}{\|\boldsymbol{x}\|_v} < \frac{1}{2}(\lambda-\rho). \qquad (4.2.65)$$

这样,取 $K>2\max\{K_0, 2c/(\lambda-\rho)\}$,

$$A = \{\boldsymbol{x}: \|\boldsymbol{x}\|_v < K\},$$

$$c_2 = \frac{1}{2}(\lambda - \rho)K - c,$$

则由(4.2.64)和(4.2.65)两式知：

$$E\{g(\boldsymbol{x}_1)\,|\,\boldsymbol{x}_0 = \boldsymbol{x}\} \leqslant \lambda\|\boldsymbol{x}\|_v - \left[\frac{1}{2}(\lambda - \rho)\|\boldsymbol{x}\|_v - c\right]$$

$$\leqslant \lambda\|\boldsymbol{x}\|_v - c_2, \quad \forall\, \boldsymbol{x} \overline{\in} A.$$

显然，由条件(iii)，存在常数 $c_3>0$，使得

$$E\{g(\boldsymbol{x}_1)\,|\,\boldsymbol{x}_0 = \boldsymbol{x}\} \leqslant c_3, \quad \forall\, \boldsymbol{x} \in A.$$

故由定理4.1.12知，模型(4.2.60)是几何遍历的.　▌

推论 4.2.19　如果定理4.2.18中的条件(ii)被推论4.2.9的条件(i)所代替，则模型(4.2.60)仍为几何遍历的.

证明　应用定理4.2.18和推论4.2.9的证明方法即可证明，具体证法从略.　▌

此推论的一种特殊情况为如下的推论：

推论 4.2.20　如果定理4.2.18的条件(ii)中的变换 φ 为线性函数，即 $\varphi(\boldsymbol{x})=\boldsymbol{\alpha}^{\tau}\boldsymbol{x}$，其中系数向量 $\boldsymbol{\alpha}=(\alpha_1,\cdots,\alpha_p)^{\tau}$ 满足条件(4.2.33)式，则模型(4.2.60)仍为几何遍历的.

定理 4.2.21　如果 NLAR 模型(4.2.60)满足下列条件：

(i) $\{\varepsilon_t\}$满足定理4.2.18的条件(i)；

(ii) φ 满足定理4.2.14的条件(4.2.48)式；

(iii) S 是正值连续函数，满足(4.2.62)式和(4.2.63)式.

则模型(4.2.60)是几何遍历的.

证明　首先，不难验证，在定理的条件下，由(4.2.61)决定的马氏链是 μ_p-不可约的、非周期的，而且任何有界 μ_p-正测集都是小集. 取准则函数

$$g(\boldsymbol{v}) = \max\{b_1|v_1|,\cdots,b_p|v_p|\}, \quad \forall\, \boldsymbol{v} \in R^p,$$

其中 $\boldsymbol{v}=(v_1,\cdots,v_p)^{\tau}$，常数 $b_1,\cdots,b_p$ 满足 $b_1>b_2>\cdots>b_p>0$，且 $\lambda b_1<b_p$，而$0<\lambda<1$由(4.2.48)式确定. 令

$$\rho = \max\{\lambda, \lambda b_1/b_{i+1}, b_{i+1}/b_i; 1 \leqslant i \leqslant p-1\}.$$

由定理的条件可得

$$E\{g(\boldsymbol{x}_1)\,|\,\boldsymbol{x}_0 = \boldsymbol{v}\} = E\{g(\boldsymbol{\varphi}(\boldsymbol{v}) + \varepsilon_1 S(\boldsymbol{v})\boldsymbol{u})\}$$

$$\leqslant E\max\{b(|\varphi(\boldsymbol{v})| + |\varepsilon_1|S(\boldsymbol{v})), b_2|v_1|,\cdots,b_p|v_{p-1}|\}$$

$$\leqslant (E|\varepsilon_1|)S(\boldsymbol{v})$$

$$+ \max\{b_1|\varphi(\boldsymbol{v})|, b_2|v_1|, \cdots, b_p|v_{p-1}|\}$$

$$\leqslant c + c'S(\boldsymbol{v})$$

$$+ \max\left\{\max\left\{\lambda b_1|v_1|, \frac{\lambda b_1}{b_2}b_2|v_2|, \cdots, \frac{\lambda b_1}{b_p}b_p|v_p|\right\},\right.$$

$$\left.\frac{b_2}{b_1}b_1|v_1|, \cdots, \frac{b_p}{b_{p-1}}b_{p-1}|v_{p-1}|\right\}$$

$$\leqslant c + c'S(\boldsymbol{v}) + \rho\max\{b_1|v_1|, \cdots, b_p|v_p|\}$$

$$\leqslant \lambda' g(\boldsymbol{v}) - [(\lambda' - \rho)g(\boldsymbol{v}) - c - c'S(\boldsymbol{v})],$$

其中$0<\rho<\lambda'<1$. 由 p 维向量范数的等价性,有

$$p^{-1/2}b_1\lambda\|\boldsymbol{v}\| \leqslant b_p\|\boldsymbol{v}\|_\infty \leqslant g(\boldsymbol{v}) \leqslant b_1\|\boldsymbol{v}\|_\infty \leqslant b_1\|\boldsymbol{v}\|.$$

由于 $\lambda'-\rho>0$,故由条件(iii),存在 $K>0$充分大及 $c_1>0$,使得当$\|\boldsymbol{v}\|\geqslant K$时,有

$$(\lambda' - \rho)g(\boldsymbol{v}) - c - c'S(\boldsymbol{x}) \geqslant c_1.$$

取 $A=\{\boldsymbol{v}: \|\boldsymbol{v}\|\leqslant K\}$作为小集,则有

$$E\{g(\boldsymbol{x}_1)|\boldsymbol{x}_0 = \boldsymbol{v}\} \leqslant \lambda' g(\boldsymbol{v}) - c_1, \quad \forall \boldsymbol{v} \overline{\in} A,$$

且

$$E\{g(\boldsymbol{x}_1)|\boldsymbol{x}_0 = \boldsymbol{v}\}$$

$$\leqslant c + c' \sup_{\|\boldsymbol{v}\|\leqslant K} S(\boldsymbol{v}) + \rho b_1 K < \infty, \quad \forall \boldsymbol{v} \in A.$$

于是由定理4.1.12知,模型(4.2.60)是几何遍历的. ▌

关于本小节的定理和推论,可参阅文献[6].

§4.3 几何遍历的 NLAR 模型的例子

在本节中,我们将利用§4.2的定理和推论,讨论第3章提出的若干非线性时间序列模型的几何遍历性问题. 若不加特别声明,本节对可加噪声 NLAR 模型类,均假定其$\{\varepsilon_t\}$为白噪声序列,并满足条件(4.2.28)和引理4.2.5的条件(ii). 对随机条件方差 NLAR 模型,均假定其$\{\varepsilon_t\}$为白噪声序列,并满足条件(3.3.19)式和引理4.2.5中条件(ii),并且 $S(\boldsymbol{x})$满足定理4.2.18的条件(iii).

例4.3.1 设时间序列$\{x_t\}$服从有界 NLAR 模型

$$x_t = \varphi(x_{t-1}, \cdots, x_{t-p}) + \varepsilon_t, \tag{4.3.1}$$

其中变换 φ 满足:对某个 $K>0$,有

$$|\varphi(\boldsymbol{x})| \leqslant K < \infty, \quad \forall \boldsymbol{x} \in R^p. \tag{4.3.2}$$

由定理4.2.8知，此类模型是几何遍历的.

类似地，若$\{x_t\}$服从随机条件方差有界 NLAR 模型

$$x_t = \varphi(x_{t-1},\cdots,x_{t-p}) + \varepsilon_t S(x_{t-1},\cdots,x_{t-p}), \qquad (4.3.3)$$

其中φ满足(4.3.2)式，则由定理4.2.18知，此类模型也是几何遍历的.

例 4.3.2 对 ANLAR 模型，即

$$x_t = c + f_1(x_{t-1}) + \cdots + f_p(x_{t-p}) + \varepsilon_t, \qquad (4.3.4)$$

如果函数f_i $(i=1,\cdots,p)$在R^1的有界集上是有界的，且存在常数$r_i>0$，$c_i\geqslant 0$，$\lambda_i>0$ $(1\leqslant i\leqslant p)$，$\lambda_1+\cdots+\lambda_p<1$，使得

$$|f_i(x_i)| \leqslant \lambda_i|x_i| + c_i, \quad |x_i| \geqslant r_i \quad (i=1,2,\cdots,p). \qquad (4.3.5)$$

不难验证，函数$\varphi(x_1,\cdots,x_p)=c+f_1(x_1)+\cdots+f_p(x_p)$在$R^p$中的有界集上有界，且存在常数$r>\max\{r_1,\cdots,r_p\}$，$c'>c+c_1+\cdots+c_p$，$\lambda=\lambda_1+\cdots+\lambda_p$，使得

$$|\varphi(x_1,\cdots,x_p)| \leqslant \lambda\max\{|x_1|,\cdots,|x_p|\} + c', \quad \forall\, \boldsymbol{x}\in R^p \text{ 且 } \|\boldsymbol{x}\| \geqslant r.$$

从而由定理4.2.14知，ANLAR 模型是几何遍历的.

例 4.3.3 对指数型 NLAR 模型(3.2.16)和(3.2.17)当$p=1$时，只要$|\varphi_{01}|<1$，由推论4.2.9知，它们都是几何遍历的. 对于模型(3.2.18)而言，只要$\boldsymbol{\alpha}=(\varphi_{01},\cdots,\varphi_{0p})^{\tau}$满足(4.2.33)式，模型(3.2.18)式即为几何遍历的.

类似地，对具有随机条件方差的指数型 NLAR 模型，例如对模型(3.2.18)，其相应的随机条件方差 NLAR 模型为(仅以$\varphi_{jk}\equiv 0$为例)：

$$x_t = \varphi_{00} + \sum_{k=1}^{p}(\varphi_{0k} + \varphi_{1k}e^{-x_{t-k}^2})x_{t-k} + \varepsilon_t S(x_{t-1},\cdots,x_{t-p}). \qquad (4.3.6)$$

只要$\boldsymbol{\alpha}=(\varphi_{01},\cdots,\varphi_{0p})^{\tau}$满足(4.2.33)式，根据推论4.2.19知，模型(4.3.6)为几何遍历的.

例 4.3.4 考虑半参数 NLAR 模型，即

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + h(x_{t-1},\cdots,x_{t-p}) + \varepsilon_t. \qquad (4.3.7)$$

如果自回归系数$\boldsymbol{\alpha}=(\alpha_1,\cdots,\alpha_p)^{\tau}$满足(4.2.33)式，函数$h$满足

$$\lim_{\|\boldsymbol{v}\|\to\infty}\frac{|h(\boldsymbol{v})|}{\|\boldsymbol{v}\|} = 0, \qquad (4.3.8)$$

$$\sup_{\|\boldsymbol{v}\|\leqslant K}|h(\boldsymbol{v})| < \infty, \quad \forall\, K>0, \qquad (4.3.9)$$

则由推论4.2.9知，此模型是几何遍历的.

类似地，在上述条件下，由推论4.2.19知，具有随机条件方差的半参数 NLAR 模型

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + h(x_{t-1}, \cdots, x_{t-p}) + \varepsilon_t S(x_{t-1}, \cdots, x_{t-p}) \tag{4.3.10}$$

也是几何遍历的.

例4.3.5 考虑 β-ARCH 模型

$$x_t = \varepsilon_t(\alpha_0 + \alpha_1 x_{t-1}^{2\beta} + \cdots + \alpha_p x_{t-p}^{2\beta})^{1/2}, \tag{4.3.11}$$

式中 $x_{t-i}^{2\beta} \equiv (x_{t-i}^2)^\beta \geqslant 0, \alpha_0 > 0, 0 \leqslant \beta \leqslant 1, \alpha_i \geqslant 0 \ (i=1,\cdots,p)$，此时还约定 $E\varepsilon_t^2 = 1$.

当 $0 \leqslant \beta < 1$ 时，易见定理4.2.18的条件均满足，从而可知模型(4.3.11)为几何遍历的.

当 $\beta = 1$ 时，(4.3.11)式为

$$x_t = \varepsilon_t(\alpha_0 + \alpha_1 x_{t-1}^2 + \cdots + \alpha_p x_{t-p}^2)^{1/2}, \tag{4.3.12}$$

此时定理4.2.18的条件(4.2.63)不成立，从而不能使用此定理. 下面直接利用定理4.1.12来证明模型(4.3.12)的几何遍历性.

定理4.3.6 若 ARCH 模型(4.3.12)中的系数满足

$$\alpha_1 + \cdots + \alpha_p < 1, \tag{4.3.13}$$

则模型(4.3.12)是几何遍历的.

证明 令 $\boldsymbol{x}_t = (x_t, \cdots, x_{t-p+1})^\tau$,

$$\boldsymbol{\varphi}(\boldsymbol{x}_{t-1}, \varepsilon_t) = (\varepsilon_t(\alpha_0 + \alpha_1 x_{t-1}^2 + \cdots + \alpha_p x_{t-p}^2)^{1/2}, x_{t-1}, \cdots, x_{t-p+1})^\tau,$$

则模型(4.3.12)转化为一阶 p 维模型

$$\boldsymbol{x}_t = \boldsymbol{\varphi}(\boldsymbol{x}_{t-1}, \varepsilon_t). \tag{4.3.14}$$

这样，为证明定理4.3.6，只需证明模型(4.3.14)是几何遍历的.

不难验证，模型(4.3.14)满足引理4.2.6的条件，从而 $\{\boldsymbol{x}_t\}$ 是不可约的、非周期的，且 R^p 中任何有界的 Lebesgue 正测集都是小集.

取准则函数

$$g(\boldsymbol{v}) = 1 + b_1 v_1^2 + b_2 v_2^2 + \cdots + b_p v_p^2, \ \forall \boldsymbol{v} \in R^p,$$

其中 $\boldsymbol{v} = (v_1, \cdots, v_p)^\tau$，系数 $b_1, b_2, \cdots, b_p$ 满足

(i) $b_1 > b_2 > \cdots > b_p > 0, \alpha_p b_1 < b_p$; (4.3.15)

(ii) $b_k < (1 - \alpha_1 - \cdots - \alpha_{k-1}) b_1, 2 \leqslant k \leqslant p$. (4.3.16)

由(4.3.14)式和 $E\varepsilon_t^2 = 1$，可得

$$
\begin{aligned}
&E\{g(\boldsymbol{x}_1)\mid \boldsymbol{x}_0=\boldsymbol{v}\}\\
&=E\left\{1+b_1\varepsilon_1^2\left(\alpha_0+\sum_{j=1}^{p}\alpha_j x_{1-j}^2\right)+\sum_{j=2}^{p}b_j x_{2-j}^2\,\middle|\,\boldsymbol{x}_0=\boldsymbol{v}\right\}\\
&=E\left\{1+b_1\varepsilon_1^2\left(\alpha_0+\sum_{j=1}^{p}\alpha_j v_j^2\right)+\sum_{j=2}^{p}b_j v_{j-1}^2\right\}\\
&\leqslant 1+b_1\alpha_0+\sum_{j=1}^{p}b_1\alpha_j v_j^2+\sum_{j=2}^{p}b_j v_{j-1}^2\\
&=c+\left(\frac{b_1\alpha_1+b_2}{b_1}\right)b_1v_1^2+\left(\frac{b_1\alpha_2+b_3}{b_2}\right)b_2v_2^2\\
&\quad+\cdots+\left(\frac{b_1\alpha_{p-1}+b_p}{b_{p-1}}\right)b_{p-1}v_{p-1}^2+\left(\frac{b_1\alpha_p}{b_p}\right)b_pv_p^2\\
&\leqslant c+\max\left\{\frac{\alpha_i b_1+b_{i+1}}{b_i},1\leqslant i\leqslant p\right\}g(\boldsymbol{v}),
\end{aligned}
\tag{4.3.17}
$$

上式中 $c=1+b_1\alpha_0, b_{p+1}=0$.

由(4.3.15)和(4.3.16)两式容易验证

$$\theta=\max\left\{\frac{\alpha_i b_1+b_{i+1}}{b_i},1\leqslant i\leqslant p\right\}<1.$$

从而存在常数 $c_1>0, 0<\theta<\lambda<1$ 和 $K>0$,使得当 $\|\boldsymbol{v}\|\geqslant K$ 时,$(\lambda-\theta)g(\boldsymbol{v})-c\geqslant c_1$. 于是取

$$A=\{\boldsymbol{v}:\|\boldsymbol{v}\|\leqslant K\},$$

则由(4.3.17)式有

$$E\{g(\boldsymbol{x}_1)\mid \boldsymbol{x}_0=\boldsymbol{v}\}\leqslant \lambda g(\boldsymbol{v})-c_1,\quad \forall\, \boldsymbol{v}\overline{\in} A,$$

$$E\{g(\boldsymbol{x}_1)\mid \boldsymbol{x}_0=\boldsymbol{v}\}\leqslant c+\theta\cdot(1+b_1K^2)<\infty,\quad \forall\, \boldsymbol{v}\in A.$$

最后,由定理4.1.12知,$\{\boldsymbol{x}_t\}$ 是几何遍历的,也即模型(4.3.12)是几何遍历的. ▌

注 4.3.7 对模型(4.3.11),作以下变换是有趣的,即在(4.3.11)式两端平方,并且取自然对数,则有

$$\log x_t^2=\log(\alpha_0+\alpha_1 e^{\beta\log x_{t-1}^2}+\cdots+\alpha_p e^{\beta\log x_{t-p}^2})+\log\varepsilon_t^2.$$

令

$$y_t=\log x_t^2,\ \ \eta_t=\log\varepsilon_t^2,$$

则上式又可写成如下的可加噪声 AR 模型:

$$y_t=\log(\alpha_0+\alpha_1 e^{\beta y_{t-1}}+\cdots+\alpha_p e^{\beta y_{t-p}})+\eta_t. \tag{4.3.18}$$

再令 $\boldsymbol{y}_t=(y_t,\cdots,y_{t-p+1})^{\tau},\ \ \boldsymbol{\xi}_t=(\eta_t,0,\cdots,0)^{\tau}$,

$$\varphi(y_{t-1}) = \log(\alpha_0 + \alpha_1 e^{\beta y_{t-1}} + \cdots + \alpha_p e^{\beta y_{t-p}}),$$

$$\boldsymbol{\varphi}(\boldsymbol{y}_{t-1}) = (\varphi(\boldsymbol{y}_{t-1}), y_{t-1}, \cdots, y_{t-p+1})^{\tau},$$

则与(4.3.18)式等价的一阶 p 维模型为

$$\boldsymbol{y}_t = \boldsymbol{\varphi}(\boldsymbol{y}_{t-1}) + \boldsymbol{\xi}_t. \tag{4.3.19}$$

易见,由(4.3.19)式确定的$\{\boldsymbol{y}_t\}$是以$(R_+^p, \mathscr{B}_p^+)$为状态空间的齐时马氏链,其中 $R_+^p = \underbrace{R_+^1 \times \cdots \times R_+^1}_{p\text{次}}$, $R_+^1 = [0, \infty)$, $\mathscr{B}_p^+$ 为 R_+^p 上的 Borel 集所成的 σ-代数. 由于对$0 \leqslant \beta \leqslant 1$,模型(4.3.18)满足引理4.2.5的条件,从而$\{\boldsymbol{y}_t\}$是不可约的、非周期的,且 R_+^p 中任何有界 Lebesgue 正测集是小集.

当$0 \leqslant \beta < 1$时,不难验证

$$|\varphi(\boldsymbol{v})| \leqslant \beta \max\{v_1, \cdots, v_p\} + c, \quad \forall\, \boldsymbol{v} \in R_+^p, \tag{4.3.20}$$

式中$\boldsymbol{v} = (v_1, \cdots, v_p)^{\tau}$. 从而由定理4.2.14知,此时模型(4.3.19)是几何遍历的.

当 $\beta = 1$时,由于

$$\varphi(\boldsymbol{v}) = \log(\alpha_0 + \alpha_1 e^{v_1} + \cdots + \alpha_p e^{v_p})$$

不满足(4.3.20)式,故不能使用定理4.2.14. 进一步,$\varphi(\boldsymbol{v})$也不满足 § 4.2 中其他定理和推论关于 $\varphi(\boldsymbol{v})$的条件,从而当 $\beta = 1$时, § 4.2的定理与推论尚不能解决(4.3.19)式的几何遍历性问题. 但是在定理4.3.6中选取适当的准则函数,已经证明,此时模型(4.3.19)是几何遍历的,在此证明中,条件(4.3.13)和 $E\varepsilon_t^2 = 1$是不可缺少的.

例4.3.8 考虑如下广义 LAR 模型

$$x_t = \varphi(\theta_1 x_{t-1} + \cdots + \theta_p x_{t-p}) + \varepsilon_t, \tag{4.3.21}$$

式中参数 $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)^{\tau}$ 满足$\|\boldsymbol{\theta}\| = 1$. 如果存在正常数 c 和 $\rho < 1$,使得

$$|\varphi(x)| \leqslant \rho |x| / \sqrt{p} + c, \tag{4.3.22}$$

于是有

$$\begin{aligned}
|\varphi(\theta_1 v_1 + \cdots + \theta_p v_p)| &\leqslant (\rho / \sqrt{p}) |\theta_1 v_1 + \cdots + \theta_p v_p| + c \\
&\leqslant \rho \max\{|v_1|, \cdots, |v_p|\} \sum_{j=1}^{p} |\theta_i| / \sqrt{p} + c \\
&\leqslant \rho \max\{|v_1|, \cdots, |v_p|\} + c.
\end{aligned}$$

这样,由定理4.2.14知,模型(4.3.21)是几何遍历的.

如果(4.3.22)式被下式代替,即存在常数 c,使 $\boldsymbol{\alpha} = c\boldsymbol{\theta}$ 满足(4.2.33)式,且

$$\lim_{|x|\to\infty}\frac{|\varphi(x)-cx|}{|x|}=0, \tag{4.3.23}$$

则由推论4.2.9知,模型(4.3.21)是几何遍历的.

类似地,在以上的条件下,具有随机条件方差的广义 LAR 模型也是几何遍历的.

例4.3.9　对分片 LAR 模型(3.2.8),即

$$x_t=\sum_{j=1}^{l}\left\{\varphi_{j0}+\sum_{k=1}^{p}\varphi_{jk}x_{t-k}\right\}I(\boldsymbol{x}_{t-1}\in R_j)+\varepsilon_t, \tag{4.3.24}$$

如果在区域 $R_1,\cdots,R_l$ 中,$R_2,\cdots,R_l$ 为有界集,而且 $\boldsymbol{\alpha}=(\varphi_{11},\cdots,\varphi_{1p})^\tau$ 满足(4.2.33)式,则容易验证定理4.2.8的条件成立,从而模型(4.3.24)是几何遍历的.

在以上条件下,随机条件方差分片 LAR 模型,即在模型(4.3.24)中用 $\varepsilon_t S(\boldsymbol{x}_{t-1})$ 代替 ε_t,也是几何遍历的.

例4.3.10　考虑门限自回归模型(3.2.9),即

$$x_t=\begin{cases}\varphi_{10}+\varphi_{11}x_{t-1}+\cdots+\varphi_{1p}x_{t-p}+\varepsilon_t,\\ \quad\text{当 } -\infty<x_{t-d}\leqslant r_2;\\ \varphi_{20}+\varphi_{21}x_{t-1}+\cdots+\varphi_{2p}x_{t-p}+\varepsilon_t,\\ \quad\text{当 } r_2<x_{t-d}\leqslant r_3;\\ \cdots\cdots\cdots\cdots\cdots\cdots\\ \varphi_{s0}+\varphi_{s1}x_{t-1}+\cdots+\varphi_{sp}x_{t-p}+\varepsilon_t,\\ \quad\text{当 } r_s\leqslant x_{t-d}<\infty.\end{cases} \tag{4.3.25}$$

如果自回归系数满足

$$\lambda=\max_{1\leqslant i\leqslant s}\sum_{j=1}^{p}|\varphi_{ij}|<1, \tag{4.3.26}$$

记

$$\varphi(v_1,\cdots,v_p)=\begin{cases}\varphi_{10}+\varphi_{11}v_1+\cdots+\varphi_{1p}v_p,\ \text{当 } -\infty<v_d\leqslant r_2;\\ \varphi_{20}+\varphi_{21}v_1+\cdots+\varphi_{2p}v_p,\ \text{当 } r_2<v_d\leqslant r_3;\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ \varphi_{s0}+\varphi_{s1}v_1+\cdots+\varphi_{sp}v_p,\ \text{当 } r_s<v_d<\infty,\end{cases}$$

容易验证

$$|\varphi(v_1,\cdots,v_p)|\leqslant\lambda\max\{|v_1|,\cdots,|v_p|\}+\max\{|\varphi_{10}|,\cdots,|\varphi_{s0}|\}.$$

上式意味着(4.2.48)式成立,故由定理4.2.14知,模型(4.3.25)是几何遍

历的.

在条件(4.3.26)下,具有随机条件方差的门限自回归模型,即在模型(4.3.25)中用 $\varepsilon_t S(\boldsymbol{x}_{t-1})$ 代替 ε_t 而得到的NLAR模型,仍是几何遍历的.

比较例4.3.9和例4.3.10可见,模型(4.3.25)是模型(4.3.24)的特殊情况. 但是,保证模型(4.3.25)为几何遍历的条件(4.3.26)式显然太强了. 其主要原因是,模型(4.3.25)中的分片区域 R_i 都是 R^p 中的无界区域,

$$R_j = \{\boldsymbol{v}:\boldsymbol{v} = (v_1,\cdots,v_p)^\tau \in R^p,\ r_j < v_d \leqslant r_{j+1}\}$$

$$(1 \leqslant j \leqslant s-1),$$

$$R_s = \{\boldsymbol{v}:\boldsymbol{v} = (v_1,\cdots,v_p)^\tau \in R^p,\ r_s < v_d < \infty\}.$$

其中 $r_1 = -\infty$.

下面的两个一阶门限自回归模型的例子表明了为保证其几何遍历性,所需的条件比(4.3.26)式弱.

例4.3.11 考虑如下一阶门限自回归模型

$$x_t = \begin{cases} \alpha_0 + \alpha_1 x_{t-1} + \varepsilon_t, & \text{当 } x_{t-1} \geqslant 0; \\ \beta_0 + \beta_1 x_{t-1} + \varepsilon_t, & \text{当 } x_{t-1} < 0, \end{cases} \tag{4.3.27}$$

其中系数 α_1 和 β_1 满足

$$\alpha_1 < 1,\ \beta_1 < 1,\ \alpha_1\beta_1 < 1. \tag{4.3.28}$$

令

$$\varphi(x) = (\alpha_0 + \alpha_1 x)I(x \geqslant 0) + (\beta_0 + \beta_1 x)I(x < 0), \tag{4.3.29}$$

则(4.3.27)式可写成

$$x_t = \varphi(x_{t-1}) + \varepsilon_t.$$

再令

$$\psi(v) = \alpha_1 v I(v \geqslant 0) + \beta_1 v I(v < 0), \tag{4.3.30}$$

容易验证

$$|\psi(v_1) - \psi(v_2)| \leqslant \max\{|\alpha_1|,|\beta_1|\}|v_1 - v_2|.$$

上式表明了 $\psi(v)$ 满足注4.2.16中的条件(4.2.57),从而 $\psi(v)$ 满足推论4.2.17中关于 $\psi(v)$ 的条件(i).

另一方面,由(4.3.28)和例2.3.22的讨论可知,ψ_k 还满足推论4.2.17中关于 $\psi(v)$ 的条件(ii). 又由(4.3.29)和(4.3.30)两式有

$$|\varphi(v) - \psi(v)| \leqslant \max\{|\alpha_0|,|\beta_0|\}.$$

此式意味着(4.2.58)式成立. 于是由推论4.2.17知,模型(4.3.27)是几何遍历的.

实际上,条件(4.3.28)还是保证模型(4.2.27)为几何遍历的必要条件(详见文献[106]).

例4.3.12　考虑如下的一阶门限自回归模型

$$x_t = \begin{cases} \alpha_{10} + \alpha_{11}x_{t-1} + \varepsilon_t, & \text{当} -\infty < x_{t-1} < r_2; \\ \alpha_{20} + \alpha_{21}x_{t-1} + \varepsilon_t, & \text{当} r_2 \leqslant x_{t-1} < r_3; \\ \cdots\cdots\cdots\cdots & \\ \alpha_{s0} + \alpha_{s1}x_{t-1} + \varepsilon_t, & \text{当} r_s \leqslant x_{t-1} < \infty, \end{cases} \tag{4.3.31}$$

其中自回归系数满足

$$\alpha_{11} < 1, \alpha_{s1} < 1, \alpha_{11} \cdot \alpha_{s1} < 1. \tag{4.3.32}$$

令　$R_1 = (-\infty, r_2]$,

$R_i = [r_i, r_{i+1})$

$(i = 2, \cdots, s;\ -\infty = r_1 < r_2 < \cdots < r_s < r_{s+1} = \infty)$,

$$\varphi(v) = \sum_{i=1}^{s} (\alpha_{i0} + \alpha_{i1}v) I(v \in R_i),$$

$$\psi(v) = \alpha_{11}vI(v < r_2) + \alpha_{s1}vI(v \geqslant r_2) + (\alpha_{s1} - \alpha_{11})r_2 I(v \geqslant r_2).$$

利用例4.3.11中类似的讨论,不难验证 $\varphi(v)$和 $\psi(v)$满足推论4.2.17的条件,从而模型(4.3.31)是几何遍历的.

例4.3.13　考虑有理分式自回归模型(3.2.20),即

$$x_t = \frac{a_0 + a_1x_{t-1} + \cdots + a_px_{t-1}^p}{b_0 + b_1x_{t-1} + \cdots + b_qx_{t-1}^q} + \varepsilon_t. \tag{4.3.33}$$

(i) 如果 $p \leqslant q$,且 $b_q \neq 0$,并存在 $\delta > 0$,使得

$$|b_0 + b_1v + \cdots + b_qv^q| \geqslant \delta > 0,$$

则

$$\varphi(v) = \frac{a_0 + a_1v + \cdots + a_pv^p}{b_0 + b_1v + \cdots + b_qv^q}$$

为有界函数,从而属于例4.3.1的特殊情况.

(ii) 如果 $p = q+1$,且 $0 < |a_p/b_q| < 1$,则取 $\alpha = a_p/b_q$,不难验证

$$\lim_{|v| \to \infty} \frac{|\varphi(v) - \alpha v|}{|v|} = 0.$$

从而仿照例4.3.8的讨论可知,此时模型(4.3.33)是几何遍历的.

例4.3.14　考虑逆自回归模型(3.4.8),即

$$f(x_t) = \alpha_0 + \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + \varepsilon_t. \qquad (4.3.34)$$

如果(4.3.34)式中的函数 f 有逆函数 $g=f^{-1}$，令 $y_t=f(x_t)$，(4.3.34)式便可写为

$$y_t = \alpha_0 + \alpha_1 g(y_{t-1}) + \cdots + \alpha_p g(y_{t-p}) + \varepsilon_t. \qquad (4.3.35)$$

如果系数 $\alpha=(\alpha_1,\cdots,\alpha_p)^\tau$ 满足(4.2.33)式，且

$$\lim_{\|v\|\to\infty} \frac{\sum_{i=1}^{p}\alpha_i^2[g(v_i)-v_i]^2}{\|v\|^2} = 0, \qquad (4.3.36)$$

则由推论4.2.9知，模型(4.3.34)是几何遍历的.

另外，若(4.3.35)式中的函数 g 为有界函数，则由例4.3.1知，模型(4.3.35)仍是几何遍历的.

例4.3.15 考虑随机条件方差门限自回归模型

$$x_t = \sum_{j=1}^{l}\left\{\varphi_{j0} + \sum_{k=1}^{p}\varphi_{jk}x_{t-k} + \varepsilon_{jt}\right\}I(x_{t-d}\in R_j), \qquad (4.3.37)$$

其中 $\{\varepsilon_{jt}\}$ 和 $R_j(1\leqslant j\leqslant l)$ 由(3.3.23)式所定义. 令

$$\varepsilon_t = \sum_{j=1}^{l}\varepsilon_{jt}I(x_{t-d}\in R_j),$$

于是模型(4.3.37)可写为

$$x_t = \sum_{j=1}^{l}\left\{\varphi_{j0} + \sum_{k=1}^{p}\varphi_{jk}x_{t-k}\right\}I(x_{t-d}\in R_j) + \varepsilon_t.$$

注意到 $\varepsilon_{jt}(1\leqslant j\leqslant l)$ 是相互独立的，用引理4.2.5的证明方法，不难验证由模型(4.3.37)产生的马氏链 $\{\boldsymbol{x}_t=(x_t,\cdots,x_{t-p+1})^\tau\}$ 是 μ_p-不可约的、非周期的，且 R^p 中有界 μ_p-正测集是小集.

再令

$$\varphi(v_1,\cdots,v_p) = \sum_{j=1}^{l}\left\{\varphi_{j0} + \sum_{k=1}^{p}\varphi_{jk}v_k\right\}I(v_d\in R_j),$$

如果条件(4.3.26)式成立，则类似于例4.3.10，存在常数 $c>0$，使得

$$|\varphi(v_1,\cdots,v_p)| \leqslant \lambda\max\{|v_1|,\cdots,|v_p|\} + c. \qquad (4.3.38)$$

从而仿照定理4.3.14的证法，可知模型(4.3.37)是几何遍历的.

§ 4.4 非线性时间序列模型的平稳解

在前两节中，我们主要讨论了 NLAR 模型的几何遍历性，包括一些常见的并在实际问题中被广泛应用的 NLAR 模型. 实际上，具有几何遍

历性或只有遍历性的非线性时间序列模型，它们必有唯一的不变概率分布，从而有唯一的平稳解. 在这一节中，我们将讨论 NLAR 模型的平稳解问题.

从前面两节的讨论可见，为了利用 § 4. 2的定理和推论来推断 NLAR 模型的几何遍历性，不可避免地要涉及由 NLAR 模型所决定的马氏链的不可约性、非周期性和小集的概念，从而导致对白噪声序列$\{\varepsilon_t\}$施加较强的限制——ε_t 有处处为正的连续密度. 在本节后面的讨论，特别是本节最后对 GARCH 模型、双线性模型和双重线性模型的平稳解的讨论，如果不强调非线性时间序列模型的遍历性，则对 ε_t 的上述过强的限制是可以被去掉的，甚至 ε_t 是否具有密度都不重要. 本节还将讨论当 ε_t 有"处处为正的连续密度"的假设不成立时的 NLAR 模型平稳解的存在问题. 与 § 4. 2和 § 4. 3的讨论一样，还将论述线性模型与非线性模型之间的差别.

一、压缩型的 NLAR 模型的平稳解

本小节讨论压缩型的 NLAR 模型平稳解的问题. 所谓压缩型的 NLAR 模型，即其相应的骨架模型中的变换 φ 满足(4. 2. 37)式的 NLAR 模型. 为便于比较线性情况与非线性情况的差别，我们再次讨论一阶 LAR 模型

$$x_t = \alpha x_{t-1} + \varepsilon_t, \tag{4.4.1}$$

其中$\{\varepsilon_t\}$为独立同分布序列，且 ε_t 与$\{x_s, s<t\}$独立，$|\alpha|<1$. 由 § 4. 2的讨论知，模型(4. 4. 1)有平稳解：

$$x_t = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j}.$$

在给定 ε_t 的分布时，模型(4. 4. 1)的不变概率分布即是

$$x_0 = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{-j}$$

的分布. 特别当 $\alpha=\frac{1}{2}$，而 ε_t 为二点分布，$P(\varepsilon_t=0)=P\left(\varepsilon_t=\frac{1}{2}\right)=\frac{1}{2}$时，由例4. 2. 12知，$x_t$ 的分布为$[0,1]$上的均匀分布，亦即$[0,1]$上的均匀分布是模型(4. 4. 1)的不变概率分布，也称为模型(4. 4. 1)的*平稳分布*. 但是对于一般情况，即使是一阶 LAR 模型，要求出其不变概率也是不容易的.

现在讨论 p 阶 LAR 模型

$$x_t = \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + \varepsilon_t, \tag{4.4.2}$$

其中$\{\varepsilon_t\}$为白噪声序列，满足(3.2.2)式，系数向量$\boldsymbol{\alpha}=(\alpha_1,\cdots,\alpha_p)^\tau$满足(4.2.33)式. 此模型的等价一阶 p 维 LAR 模型为(4.2.40)式. 如本章§4.1所叙述的那样，满足(4.2.40)式的$\{\boldsymbol{x}_t\}$与满足(4.4.2)的$\{x_t\}$一一对应，而$\{\boldsymbol{x}_t\}$的概率结构被(4.2.40)的转移概率和$\boldsymbol{x}_0$的初始分布所唯一确定. 必须注意，(4.2.40)的不变概率分布是 p 维分布，而且保证

$$\boldsymbol{x}_0 = (x_0, x_{-1}, \cdots, x_{-p+1})$$

与

$$\boldsymbol{x}_1 = (\alpha_1 x_0 + \alpha_2 x_{-1} + \cdots + \alpha_p x_{-p+1} + \varepsilon_1, x_0, \cdots, x_{-p+2})$$

具有相同的分布. 换言之，LAR(p)模型(4.4.2)的不变概率分布是指使得$(x_t, x_{t-1}, \cdots, x_{t-p+1})$与$(x_{t-1}, x_{t-2}, \cdots, x_{t-p})$具有相同分布的 p 维分布，不要误以为仅指 x_t 与 x_{t-1}具有相同分布的一维分布.

利用第2章中的矩阵约当标准型方法，可知级数

$$\boldsymbol{x}_t = \sum_{j=0}^{\infty} A^j \varepsilon_{t-j} \boldsymbol{u} \tag{4.4.3}$$

是依概率收敛的，而且容易验证它满足(4.2.40)式，即$\{\boldsymbol{x}_t\}$是(4.2.40)式的平稳解，$\boldsymbol{x}_0$的分布为(4.2.40)式的不变概率分布. 其唯一性也不难验证，在此从略.

对于 LAR(p)模型(4.4.2)而言，保证它有唯一平稳解的关键条件是：系数向量$\boldsymbol{\alpha}$满足(4.2.33)式，从而保证(4.2.40)式的骨架模型

$$\begin{cases} \boldsymbol{y}_t = A\boldsymbol{y}_{t-1}; \\ \boldsymbol{y}_0 \in R^p, \end{cases} \tag{4.4.4}$$

按某范数具有压缩性(见定理2.3.9). 由一阶 p 维 LAR 模型(4.2.40)的平稳解(4.4.3)式，不难求得模型(4.4.2)的平稳解. 在这一求解过程中，只要$\{\varepsilon_t\}$能保证(4.4.3)式的级数在依概率意义下收敛即可，并不要求 ε_t 的分布具有正密度. 实际上，能使(4.4.3)式依概率收敛的充分必要条件是

$$E\max\{1, \log|\varepsilon_t|\} < \infty \tag{4.4.5}$$

(见文献[1]). 还需指出，虽然对 LAR 模型的平稳解存在性的讨论是很容易的事，但是要求出其平稳分布的具体形式却不容易，除非对某些极特殊的情况——如正态情况(见文献[3]).

对于 NLAR 模型，求其平稳解和不变概率分布就更加困难了，例如对可加噪声 NLAR 模型(4.2.27)及其等价的一阶 p 维模型(4.2.29)，一

般说来，很难求出其平稳解的明显表达式，而只能用类似于(1.5.8)式的形式表达式，即应用(4.2.29)式迭代 t 步：

$$\begin{aligned} \boldsymbol{x}_t &= \boldsymbol{\varphi}\,(\boldsymbol{x}_{t-1}) + \varepsilon_t \boldsymbol{u} \\ &= \boldsymbol{\varphi}\,(\boldsymbol{\varphi}\,(\boldsymbol{x}_{t-2}) + \varepsilon_{t-1}\boldsymbol{u}) + \varepsilon_t \boldsymbol{u} \\ &= \boldsymbol{\varphi}\,(\boldsymbol{\varphi}\,(\boldsymbol{\varphi}\,(\boldsymbol{x}_{t-3}) + \varepsilon_{t-2}\boldsymbol{u}) + \varepsilon_{t-1}\boldsymbol{u}) + \varepsilon_t \boldsymbol{u} \\ &= \cdots\cdots\cdots \\ &= \boldsymbol{\varphi}\,(\boldsymbol{\varphi}\,(\boldsymbol{\varphi}\,(\cdots(\boldsymbol{\varphi}\,(\boldsymbol{\varphi}\,(\boldsymbol{x}_0) + \varepsilon_1\boldsymbol{u}) + \varepsilon_2\boldsymbol{u})\cdots) \\ &\quad + \varepsilon_{t-2}\boldsymbol{u}) + \varepsilon_{t-1}\boldsymbol{u}) \\ &\quad + \varepsilon_t \boldsymbol{u}. \end{aligned} \tag{4.4.6}$$

当然，(4.4.6)式并不是模型(4.2.27)的解的表达式. 不过根据§4.2的定理和推论，当模型(4.2.29)的骨架模型和$\{\varepsilon_t\}$满足适当条件时，例如条件(4.2.31)和定理4.2.8的条件(ii)，可以保证模型(4.2.29)有唯一的平稳解. 现在我们讨论在(4.2.29)式的骨架模型(4.2.30)具有适当的压缩性质时，放宽定理4.2.8的条件(ii)关于 ε_t 的要求，仍能保证模型(4.2.29)有唯一的平稳解. 其基本技巧在§1.5中曾提到过，即在特殊情况下，可使用后移法证明一阶一维 NLAR 模型的平稳解的存在性，并提出如何给出平稳解的逼近. 将这一方法推广到模型(4.2.29)，便可得到如下定理：

定理 4.4.1　若 NLAR 模型(4.2.29)的骨架模型(4.2.30)满足条件

$$\|\boldsymbol{\varphi}\,(\boldsymbol{x}) - \boldsymbol{\varphi}\,(\boldsymbol{y})\|_v \leqslant \rho\|\boldsymbol{x} - \boldsymbol{y}\|_v,\quad \forall\ \boldsymbol{x},\boldsymbol{y} \in R^p, \tag{4.4.7}$$

其中$\|\cdot\|_v$为某 p 维向量范数，正数 $\rho<1$，$\{\varepsilon_t\}$为白噪声序列并且满足(3.2.2)式，则模型(4.2.29)有唯一的平稳解.

证明　由(4.4.7)式，取 $\boldsymbol{y}=\boldsymbol{0}$，可得

$$\|\boldsymbol{\varphi}\,(\boldsymbol{x}) - \boldsymbol{\varphi}\,(\boldsymbol{0})\|_v \leqslant \rho\|\boldsymbol{x}\|_v,\quad \forall\ \boldsymbol{x} \in R^p. \tag{4.4.8}$$

若以 $\tilde{\boldsymbol{\varphi}}(\boldsymbol{x})=\boldsymbol{\varphi}\,(\boldsymbol{x})-\boldsymbol{\varphi}\,(\boldsymbol{0})$代替 $\boldsymbol{\varphi}\,(\boldsymbol{x})$，则 $\tilde{\boldsymbol{\varphi}}(\boldsymbol{0})=\boldsymbol{0}$，所以不妨假定 $\boldsymbol{\varphi}\,(\boldsymbol{0})=\boldsymbol{0}$. 于是代替(4.4.8)式，我们使用

$$\|\boldsymbol{\Phi}\,(\boldsymbol{x})\|_v \leqslant \rho\|\boldsymbol{x}\|_v,\quad \forall\ \boldsymbol{x} \in R^p, \tag{4.4.9}$$

此式表明 $\boldsymbol{\varphi}$ 具有压缩性质. 沿用(4.4.6)式的记号，令

$$\xi_{n,t} = \varepsilon_t\boldsymbol{u} + \boldsymbol{\Phi}\,(\boldsymbol{\Phi}\,(\cdots\boldsymbol{\varphi}\,(\boldsymbol{\varphi}\,(\boldsymbol{o}) + \varepsilon_{t-n+1}\boldsymbol{u}) + \varepsilon_{t-n+2}\boldsymbol{u})\cdots) + \varepsilon_{t-1}\boldsymbol{u}). \tag{4.4.10}$$

上式中的 $\xi_{n,t}$表示从 $\boldsymbol{x}_{t-n}=\boldsymbol{0}$出发用(4.2.29)式迭代 n 次的取值. 用类似于§1.5的推理方法，可得

$$\|\xi_{n,t} - \xi_{m,t}\|_v \leqslant \rho\|\xi_{n-1,t-1} - \xi_{m-1,t-1}\|_v$$

$$
\begin{aligned}
&\leqslant \cdots\cdots \\
&\leqslant \rho^{m-1}\|\boldsymbol{\xi}_{n-m+1,t-m+1} - \boldsymbol{\xi}_{1,t-m+1}\|_v \\
&\leqslant \rho^{m-1}\|\boldsymbol{\varphi}(\boldsymbol{\xi}_{n-m,t-m}) + \varepsilon_{t-m+1} - \varepsilon_{t-m+1}\|_v \\
&\leqslant \rho^{m-1}\|\boldsymbol{\varphi}(\boldsymbol{\xi}_{n-m,t-m}) - \boldsymbol{\varphi}(\mathbf{0})\|_v \\
&\leqslant \rho^{m}\|\boldsymbol{\xi}_{n-m,t-m}\|_v \\
&= \rho^{m}\|\boldsymbol{\varphi}(\boldsymbol{\xi}_{n-m-1,t-m-1}) + \varepsilon_{t-m}\boldsymbol{u}\|_v \\
&\leqslant \rho^{m}\{|\varepsilon_{t-m}|\cdot\|\boldsymbol{u}\|_v + \rho\|\boldsymbol{\xi}_{n-m-1,t-m-1}\|_v\} \\
&\leqslant \rho^{m}\{\delta_1|\varepsilon_{t-m}| + \rho\delta_1|\varepsilon_{t-m-1}| + \rho^2\|\boldsymbol{\xi}_{n-m-2,t-m-2}\|_v\} \\
&\leqslant \cdots\cdots \\
&\leqslant \rho^{m}\{\delta_1|\varepsilon_{t-m}| + \rho\delta_1|\varepsilon_{t-m-1}| + \cdots + \rho^{n-m}\delta_1|\varepsilon_{t-n}|\} \\
&\leqslant \rho^{m}\delta_1\{|\varepsilon_{t-m}| + \rho|\varepsilon_{t-m-1}| + \cdots + \rho^{n-m}|\varepsilon_{t-n}| + \cdots\} \\
&= \rho^{m}\delta_1\sum_{j=0}^{\infty}\rho^{j}|\varepsilon_{t-m-j}|,
\end{aligned}
\tag{4.4.11}
$$

其中 $\delta_1 = \|\boldsymbol{u}\|_v$. 由于(4.4.11)式右端的随机级数是依概率收敛的,从而

$$\|\boldsymbol{\xi}_{n,t} - \boldsymbol{\xi}_{m,t}\|_v \to 0,\ \text{a.s.},\ 当\ n、m \to \infty.$$

故$\{\boldsymbol{\xi}_{n,t}\}$为 Cauchy 序列. 于是$\{\boldsymbol{\xi}_{n,t}\}$概率为1地有极限,记为

$$\boldsymbol{x}_t = \lim_{n\to\infty}\boldsymbol{\xi}_{n,t},\ \text{a.s.}. \tag{4.4.12}$$

再由(4.2.29)式和(4.4.10)式,有

$$\boldsymbol{\xi}_{n,t} = \varphi(\boldsymbol{\xi}_{n-1,t-1}) + \varepsilon_t\boldsymbol{u}.$$

又由条件(4.4.7)式知,$\boldsymbol{\varphi}$ 为连续的,从而由上式和(4.4.12)式,有

$$\boldsymbol{x}_t = \boldsymbol{\varphi}(\boldsymbol{x}_{t-1}) + \varepsilon_t\boldsymbol{u}.$$

由(4.4.10)式和(4.4.12)式知,$\{\boldsymbol{x}_t\}$为平稳的,即$\{\boldsymbol{x}_t\}$是(4.2.29)式的平稳解,解的唯一性是显然的. ▌

在定理4.4.1中,我们只保持了对$\{\varepsilon_t\}$的最基本的假定,即$\{\varepsilon_t\}$为白噪声序列,且满足(3.2.2)式.如果假定 ε_t 的分布满足定理4.2.8的条件(ii),而(4.4.8)式保证了定理4.2.8的条件(i),则直接由定理4.2.8知,模型(4.2.29)有平稳解.但在定理4.4.1的证明中,既没有使用关于 ε_t 的较强的条件——定理4.2.8的条件(ii),也没有涉及马氏链的概念.事实上,定理4.4.1中对$\{\varepsilon_t\}$的假定还可以放宽.因为在定理4.4.1的证明中,需要用到(4.4.11)式右端随机级数依概率收敛的性质,所以根据$\{\varepsilon_t\}$独立同分布的假定,只要条件(4.4.5)式满足,便有(4.4.11)式右端随机级数依概率收敛.这一条件不能再改进了,除非 $\boldsymbol{\varphi}(x)\equiv\mathbf{0}$.

由上述讨论可见，使用§4.2中的定理或推论来判断压缩型的NLAR模型有无平稳解，似乎有点“大材小用”，反而对$\{\varepsilon_t\}$施加了过强的限制，而定理4.4.1放宽了这一限制. 人们自然会想到，能否进一步推广定理4.4.1呢？例如将条件(4.4.7)放宽为§4.2中诸定理的条件？我们不能简单地回答这一问题. 正如在§4.2中注4.2.13中所论述的那样，在条件(4.4.7)下，模型(4.2.29)的骨架模型(4.2.30)的吸引集是单点集，并且是整体稳定的. 这与平稳线性情况很相似. 而在§4.2的多数定理和推论中，相应的骨架模型(4.2.30)的吸引集都是有界集. 这与定理4.4.1的条件有本质区别.

最后，我们还要指出，在§4.1中所定义的不可约概念在讨论NLAR模型的平稳解问题时，也有其不便之处. 仍以模型(4.4.1)为例，取$\alpha=\frac{1}{2}$，即

$$x_t = \frac{1}{2}x_{t-1} + \varepsilon_t, \tag{4.4.13}$$

其中$\{\varepsilon_t\}$为白噪声序列，服从二点分布$P(\varepsilon_t=1)=P(\varepsilon_t=-1)=\frac{1}{2}$. 如前所述，模型(4.4.13)有平稳解

$$x_t = \sum_{j=0}^{\infty}\left(\frac{1}{2}\right)^j \varepsilon_{t-j}. \tag{4.4.14}$$

又如在注4.2.11中所讨论的那样，若以$\|F-G\|_\infty$代替全变差范数，则模型(4.4.13)还具有几何遍历性. 现在还要指出，按§4.1的不可约定义，由(4.4.13)式决定的马氏链不是不可约的. 因为当初始值x_0为有理数时，显然$x_1,x_2,\cdots$都是有理数. 于是记$A=\{x:x$ 为无理数$\}$，必有

$$P(x,A)=0,\quad \forall\, \boldsymbol{x} \overline{\in} A. \tag{4.4.15}$$

不仅如此，对任意正整数k，$P_k(x,A)=0$，而且$\mu(A^c)=0$，这里μ为Lebesgue 测度. 另一方面，对于任意$y\in A$，从y出发，由(4.4.13)迭代产生的$x_1,x_2,\cdots$均为无理数，从而

$$P(y,A^c)=0. \tag{4.4.16}$$

联合(4.4.15)和(4.4.16)两式可知，由(4.4.13)决定的马氏链不可能是不可约的. 由于这样，无法使用§4.2中的任何定理和推论来判断模型(4.4.13)有无平稳解. 但是(4.4.14)式表明了模型(4.4.13)确有唯一的平稳解，而且如注解4.2.11中所述，在某种意义下，模型(4.4.13)还是几何遍历的.

综上所述,从研究NLAR模型的平稳解和平稳分布的角度来看,使用§4.1中的马氏链的漂移准则的方法似有不足之处.

二、非压缩型的NLAR模型的平稳解

在定理4.4.1中,被讨论的NLAR模型(4.2.29)中的变换 φ 具有压缩性(见(4.4.7)式),从而相应的骨架模型(4.2.30)的吸引集为单点集,且为指数速度整体稳定的.现在讨论不具有压缩性的NLAR模型有平稳解的条件问题.为了论述方便,以下所述压缩性均指模型的变换满足(4.4.7)式.当NLAR模型(4.2.29)不具有压缩性时,一般说来,相应的骨架模型(4.2.30)的吸引集不仅仅是含有单点元素,这时定理4.4.1不再适用了,因此,其证明方法也不再奏效.下面先分析一个例子.

例4.4.2 考虑如下分段LAR模型:

$$x_t = \begin{cases} \dfrac{1}{2} + \dfrac{1}{2}x_{t-1} + \varepsilon_t, & \text{当 } x_{t-1} \geqslant 0; \\ -\dfrac{1}{2} + \dfrac{1}{2}x_{t-1} + \varepsilon_t, & \text{当 } x_{t-1} < 0, \end{cases} \tag{4.4.17}$$

其中 $\{\varepsilon_t\}$ 为白噪声序列,满足(4.2.28)式.如果 $\{\varepsilon_t\}$ 还满足定理4.2.8的条件,则模型(4.4.17)是几何遍历的.从而不论 x_0 为何值,x_t 的分布都以指数速度收敛于模型(4.4.17)的不变概率分布.但是,如果 ε_t 不满足定理4.2.8的条件,模型(4.4.17)就不一定是几何遍历的.比如当 ε_t 服从二点分布,且 $P\left(\varepsilon_t=\dfrac{1}{4}\right)=P\left(\varepsilon_t=-\dfrac{1}{4}\right)=\dfrac{1}{2}$ 时,模型(4.4.17)的不变概率分布就不唯一(见例4.2.3),从而平稳解也不唯一.所以模型(4.4.17)不可能具有遍历性.

上述现象与模型(4.4.17)的骨架模型有关,其骨架模型为

$$y_t = \begin{cases} \dfrac{1}{2} + \dfrac{1}{2}y_{t-1}, & \text{当 } y_{t-1} \geqslant 0; \\ -\dfrac{1}{2} + \dfrac{1}{2}y_{t-1}, & \text{当 } y_{t-1} < 0. \end{cases} \tag{4.4.18}$$

易见,(4.4.18)式有两个不动点 $\{-1\}$ 和 $\{1\}$,而且它们的吸引域分别是 $[0,\infty)$ 和 $(-\infty,0)$.它们都是局部几何速度稳定的不动点.显然,当 ε_t 的取值满足 $|\varepsilon_t|<½$ 时,若(4.4.17)式从初始值 $x_0\in[0,\infty)$ 出发,产生的迭代序列 $\{x_t\}$ 仍满足 $x_t\in[0,\infty)$,于是 x_t 的分布收敛到

$$u_t = 1 + \sum_{j=0}^{\infty}\left(\frac{1}{2}\right)^j \varepsilon_{t-j}$$

的分布. 反之,若从初始值 $x_0 \in (-\infty, 0)$ 出发,类似地有 $x_t \in (-\infty, 0)$,从而 x_t 的分布收敛到

$$v_t = -1 + \sum_{j=0}^{\infty}\left(\frac{1}{2}\right)^j \varepsilon_{t-j}$$

的分布. 由此可见,对骨架模型(4.4.18)增添了随机化的"外衣"ε_t 以后的(4.4.17)式并不能将两个不同的吸引元$\{1\}$和$\{-1\}$联接起来,或形象地说,这件随机化"外衣"还太"单薄",不能把骨架完全"遮隐"起来. 如果将随机化的"外衣"ε_t 加强到满足定理4.2.8的条件(ii),则由该定理可知,模型(4.4.17)有唯一的不变概率分布,而且具有遍历性. 至此,人们自然会问,若只为保证模型(4.4.17)有唯一平稳解,那么 ε_t 的分布的最低限制是什么呢?根据§4.2的定理及其证明过程可见,至少应当保证(4.4.17)式所决定的马氏链是不可约的. 为了使$\{x_t\}$是不可约的,至少要求

$$\limsup_{t\to\infty} P(x_t \in [0,\infty) \mid x_0 = x) > 0, \quad \forall\, x \in R^1, \tag{4.4.19}$$

$$\limsup_{t\to\infty} P(x_t \in (-\infty,0) \mid x_0 = x) > 0, \quad \forall\, x \in R^1. \tag{4.4.20}$$

由上述讨论不难看出,(4.4.19)和(4.4.20)两式成立的充分必要条件是

$$P\left(\varepsilon_t > \frac{1}{2}\right) > 0,\ P\left(\varepsilon_t < -\frac{1}{2}\right) > 0. \tag{4.4.21}$$

综上所述,当 NLAR 模型(4.2.29)中的 φ 不满足压缩性,而满足§4.2中的某些定理或推论的条件时,模型(4.2.29)是否有唯一平稳解和不变概率分布,还取决于 ε_t 的分布性质.

在 φ 不具有压缩性时,不能用证明定理4.4.1中的方法构造柯西序列去逼近 NLAR 模型的平稳解. 进一步,若 NLAR 模型(4.2.29)的骨架模型中的变换 φ 满足定理4.2.8的条件(i),但 ε_t 不满足该定理的条件(ii),我们既不能利用定理4.4.1,也不能用定理4.2.8来判断模型是否有平稳解. 下面讨论在此情形下,NLAR 模型平稳解的存在问题. 先讨论如下的分片取常数值的自回归模型

$$x_t = \sum_{j=1}^{l} a_j I(x_{t-1} \in R_j) + \varepsilon_t, \tag{4.4.22}$$

式中 $a_1, \cdots, a_l$ 为常数,$R_1, \cdots, R_l$ 为 R^1的一个有限分割. 实际上,(4.4.22)式是(3.2.8)式的特例. 显然,如果$\{\varepsilon_t\}$是白噪声序列,满足条件(4.2.28)

式和定理4.2.8的条件(ii)，那么模型(4.2.22)是几何遍历的. 当$\{\varepsilon_t\}$不满足定理4.2.8的条件(ii)时，模型(4.4.22)是否还有平稳解呢？下面通过求模型(4.4.22)的不变概率分布来解决这一问题. 记

$$\varphi(x) = \sum_{j=1}^{l} a_j I(x \in R_j), \quad q_j = P(x_{t-1} \in R_j).$$

注意到ε_t与x_{t-1}独立，由(4.2.22)式有

$$\begin{aligned} P(x_t < x) &= P(\varphi(x_{t-1}) + \varepsilon_t < x) \\ &= \sum_{j=1}^{l} P(\varphi(x_{t-1}) + \varepsilon_t < x, x_{t-1} \in R_j) \\ &= \sum_{j=1}^{l} P(a_j + \varepsilon_t < x) P(x_{t-1} \in R_j) \\ &= \sum_{j=1}^{l} F_\varepsilon(x - a_j) q_j, \quad \forall\, x \in R^1, \end{aligned} \tag{4.4.23}$$

式中F_ε为ε_t的分布函数. 在(4.4.23)式中，当给定x时，$F_\varepsilon(x-a_j)$是已知的. 因此，为求x_t的不变分布$P(x_t<x)$，只需求出概率值$q_j(j=1,\cdots,l)$即可. 再由(4.4.22)式、ε_t与x_{t-1}独立及ε_t的同分布性质，又有

$$\begin{aligned} P(x_t \in R_k) &= P(\varphi(x_{t-1}) + \varepsilon_t \in R_k) \\ &= \sum_{i=1}^{l} P(a_i + \varepsilon_1 \in R_k) P(x_{t-1} \in R_i). \end{aligned} \tag{4.4.24}$$

记

$$p_{ik} = P(a_i + \varepsilon_1 \in R_k),$$

由(4.4.24)知，模型(4.4.22)的不变概率分布应满足如下方程：

$$q_k = \sum_{i=1}^{l} p_{ik} q_i \quad (k = 1, 2, \cdots, l). \tag{4.4.25}$$

记

$$\boldsymbol{q} = (q_1, \cdots, q_l)^\tau,$$

$$P = \begin{pmatrix} p_{11} & p_{21} & \cdots & p_{l1} \\ p_{12} & p_{22} & \cdots & p_{l2} \\ \cdots & \cdots & \cdots & \cdots \\ p_{1l} & p_{2l} & \cdots & p_{ll} \end{pmatrix},$$

于是，(4.4.25)式可写成如下矩阵方程：

$$\boldsymbol{q} = P\boldsymbol{q}. \tag{4.4.26}$$

以下证明，在一定的条件下，矩阵方程(4.4.26)存在唯一的非负解. 为此，

引进如下记号和定义，对矩阵 $A=(a_{ij})_{l\times l}$ 和向量 $\boldsymbol{b}=(b_1,\cdots,b_l)^\tau$，记

$$A\gg 0 \ \text{表示}\ a_{ij}\geqslant 0 \quad (i,j=1,\cdots,l);$$

$$\boldsymbol{b}\gg \boldsymbol{0} \ \text{表示}\ b_i\geqslant 0 \quad (i=1,\cdots,l);$$

$$P\boldsymbol{b}\gg \lambda\boldsymbol{b} \ \text{表示}\ P\boldsymbol{b}-\lambda\boldsymbol{b}\gg \boldsymbol{0}.$$

再记

$\Lambda(A)=\{\lambda$：存在向量 $\boldsymbol{b}\gg\boldsymbol{0}$，使得 $A\boldsymbol{b}\geqslant\lambda\boldsymbol{b}$，且 $b_1+\cdots+b_l=1\}$，

$\lambda_0=\sup\{\lambda:\lambda\in\Lambda(A)\}$.

引理 4.4.3　如果矩阵 $A\gg0$，且 λ_0 是 A 的一个特征值，则必存在向量 $\boldsymbol{b}_0\gg\boldsymbol{0}$，使得 $A\boldsymbol{b}_0=\lambda_0\boldsymbol{b}_0$.

证明见文献[68].

引理 4.4.4　矩阵方程(4.4.26)式必有非负解.

证明　由矩阵 P 的定义知

$$p_{ik}\geqslant 0,\quad p_{i1}+\cdots+p_{il}=1,\ (i=1,\cdots,l),\qquad (4.4.27)$$

即矩阵 P 是一转移矩阵. 记 $\boldsymbol{e}=(1,1,\cdots,1)^\tau$，则有

$$P^\tau\boldsymbol{e}=\boldsymbol{e},$$

即 1 是 P^τ 的特征值，从而也是 P 的特征值. 又由 $\Lambda(P)$ 的定义知，$1\in\Lambda(P)$，且对任意的 $\lambda\in\Lambda(P)$，必有 $\boldsymbol{b}\gg0$，使得

$$P\boldsymbol{b}\gg\lambda\boldsymbol{b}.$$

将上式两端向量的元素各自相加，由(4.4.27)式有

$$\sum_{k=1}^{l}\sum_{i=1}^{l}p_{ik}b_i=\sum_{i=1}^{l}\left(\sum_{k=1}^{l}p_{ik}\right)b_i=1$$

$$\geqslant\lambda\sum_{i=1}^{l}b_i=\lambda.$$

从而由 λ_0 的定义知 $\lambda_0=1$. 再由引理4.4.3便得引理的结论.　▌

引理4.4.5　如果存在正整数 $k\geqslant1$ 及 $1\leqslant s\leqslant l$，使得

$$P^k=(p_{ij}^{(k)})_{1\leqslant i,j\leqslant l}$$

中的第 s 行元素

$$p_{sj}^{(k)}>0\quad(j=1,2,\cdots,l),\qquad(4.4.28)$$

则矩阵方程(4.4.26)的解是唯一的.

证明　由文献[43]和条件(4.4.28)式知

$$\lim_{n\to\infty}p_{sj}^{(n)}=b_j\geqslant0,\ \text{且}\sum_{j=1}^{l}b_j=1.\qquad(4.4.29)$$

令 $\boldsymbol{b}=(b_1,\cdots,b_l)^\tau$，$Q=(\boldsymbol{b},\boldsymbol{b},\cdots,\boldsymbol{b})$，则(4.4.29)式意味着

$$\lim_{n\to\infty} P^n Q = Q,\quad PQ = Q.$$

从而

$$P\boldsymbol{b} = \boldsymbol{b}.$$

由(4.4.26)式知 $\boldsymbol{q}=\boldsymbol{b}$，且被唯一决定. ▌

由引理4.4.4和引理4.4.5知，如果条件(4.4.28)成立，则矩阵方程(4.4.26)有唯一的非负解，从而由(4.4.23)和(4.4.24)两式知，模型(4.4.22)有唯一的不变概率分布. 虽然寻找使(4.4.28)式成立的充分必要条件不容易，但是，在很多情况下，其充分条件却是容易给出的. 例如当 $R_1,\cdots,R_l$ 为 l 个区间

$$R_j = (r_{j-1}, r_j] \quad (j = 1,2,\cdots,l-1),\quad R_l = (r_{l-1},\infty),$$

其中 $-\infty=r_0<r_1<\cdots<r_{l-1}<r_l=\infty$ ，记

$$a^- = \min(a_1,\cdots,a_l),\quad a^+ = \max(a_1,\cdots,a_l).$$

如果 ε_t 的分布为离散分布，且落在$[a^-+r_l, a^++r_{l-1}]$中的每个有理数上的概率大于0，易见(4.4.25)式中的每个 $p_{ik}>0$，从而引理4.4.5中的条件(4.4.28)式成立，且相应的 $k=1, s=1$（实际上 s 可取1至 l 之间的任何整数），于是模型(4.4.22)有唯一的平稳解.

特别地，当模型(4.4.22)有如下形式：

$$x_t = \begin{cases} a_1 + \varepsilon_t, & \text{当 } x_{t-1} \geqslant 0; \\ a_2 + \varepsilon_t, & \text{当 } x_{t-1} < 0, \end{cases} \tag{4.4.30}$$

式中 $a_1>0$，$a_2<0$. 如果 ε_t 的分布满足

$$P(\varepsilon_t < -a_1) > 0 \quad P(\varepsilon_t \geqslant -a_2) > 0, \tag{4.4.31}$$

则引理4.4.5的条件(4.4.28)成立，从而模型(4.4.30)有唯一的平稳解.
例如，若 ε_t 的分布为

$$\begin{cases} P(\varepsilon_t = -a_1 - \delta) = p > 0; \\ P(\varepsilon_t = -a_2 + \delta) = 1 - p > 0, \end{cases} \tag{4.4.32}$$

式中 δ 为任意正实数，则容易验证(4.4.31)式成立. 这时模型(4.4.30)的不变概率分布可由(4.4.23)式和(4.4.24)式给出. 例如取

$$\delta > \max\{a_1, -a_2\},$$

则
$$P(a_1 + \varepsilon_t \geqslant 0) = P(\varepsilon_t = -a_2 + \delta) = 1 - p \triangleq p_{11},$$
$$P(a_1 + \varepsilon_t < 0) = P(\varepsilon_t = -a_1 - \delta) = p \triangleq p_{12},$$
$$P(a_2 + \varepsilon_t \geqslant 0) = P(\varepsilon_t = -a_2 + \delta) = 1 - p \triangleq p_{21},$$
$$P(a_2 + \varepsilon_t < 0) = P(\varepsilon_t = -a_1 - \delta) = p \triangleq p_{22}.$$

将上述四式代入(4.4.24)式,得

$$q_1 = 1 - p,\ q_2 = p.$$

再将 q_1、q_2代入(4.4.23)式,得

$$\begin{aligned} F(x) &= P(x_t < x) = F_\varepsilon(x - a_1)q_1 + F_\varepsilon(x - a_2)q_2 \\ &= F_\varepsilon(x - a_1)(1 - p) + F_\varepsilon(x - a_2)p \\ &= \begin{cases} 0, & \text{当 } x < a_2 - a_1 - \delta; \\ p^2, & \text{当 } a_2 - a_1 - \delta \leqslant x < -\delta; \\ p, & \text{当 } -\delta \leqslant x < \delta; \\ 2p - p^2, & \text{当 } \delta \leqslant x < a_1 - a_2 + \delta; \\ 1, & \text{当 } x \geqslant a_1 - a_2 + \delta. \end{cases} \end{aligned} \tag{4.4.33}$$

另外,通过直接验证可知,无论(4.4.30)式中的初始值 $x_0 = a$ 取什么值,x_2的条件分布 $P(x_2 < x | x_0 = a)$都与(4.4.33)式给出的不变概率分布相等. 进一步还可验证 $x_3, x_4, \cdots$ 的条件分布也是如此. 于是,由定义4.2.4知,如果 ε_t 服从由(4.4.32)式给出的离散分布,则模型(4.4.30)是几何遍历的.

但讨论模型(4.4.22)有唯一平稳解的方法,对于一般的NLAR(p)模型却不适用,具体理由如下:

考虑分片取常数值的 NLAR(p)模型

$$x_t = \sum_{j=1}^{l} a_j I(\boldsymbol{x}_{t-1} \in R_j) + \varepsilon_t, \tag{4.4.34}$$

其中 $R_j (j = 1, \cdots, l)$是 R^p 中的有限分割,$\{\varepsilon_t\}$是白噪声序列,满足(3.2.2)式. 将(4.4.34)式改写成相应的一阶 p 维 NLAR 形式,再使用(4.4.24)式的推导方法,并不能得出(4.4.25)式. 其原因在于,事件$\{\boldsymbol{\varphi}(\boldsymbol{x}_{t-1}) + \varepsilon_t \boldsymbol{u} \in R_k\}$与$\{\boldsymbol{x}_{t-1} \in R_j\}$不再相互独立,其中

$$\boldsymbol{\varphi}(\boldsymbol{x}_{t-1}) = (a_l I(\boldsymbol{x}_{t-1} \in R_1) + \cdots + a_1 I(\boldsymbol{x}_{t-1} \in R_l), x_{t-1}, \cdots, x_{t-p+1})^\tau,$$

也即

$$\begin{aligned} &P(\boldsymbol{\varphi}(\boldsymbol{x}_{t-1}) + \varepsilon_t \boldsymbol{u} \in R_k, \boldsymbol{x}_{t-1} \in R_j) \\ &\quad = P((a_j + \varepsilon_t, x_{t-1}, \cdots, x_{t-p+2})^\tau \in R_k, \boldsymbol{x}_{t-1} \in R_j) \\ &\quad \neq P((a_j + \varepsilon_t, x_{t-1}, \cdots, x_{t-p+2})^\tau \in R_k) P(\boldsymbol{x}_{t-1} \in R_j), \end{aligned}$$

只有当 $p=1$时,在(4.4.24)式中才能有上述等式成立.

另外,当 $p=1$时,可用讨论模型(4.2.22)有唯一平稳解的方法解决$\varphi(x)$为有界连续函数的 NLAR(1)模型的平稳解问题. 若 NLAR(1)模型中的 φ 为有界连续函数,考虑如下的近似模型

$$x_t = \sum_{j=1}^{l} a_j I(\varphi(x_{t-1}) \in R_j) + \varepsilon_t, \tag{4.4.35}$$

式中

$$R_1 = (-\infty, r_1),$$
$$R_{i+1} = [r_i, r_{i+1})$$
$$(i = 1, \cdots, l-1; -\infty = r_0 < r_1 < \cdots < r_{l-1} = r_l = \infty),$$

$\{\varepsilon_t\}$为白噪声序列,满足(3.2.2)式. 近似模型(4.4.35)可按照模型(4.4.22)的方法求出其不变概率分布. 当(4.4.35)式中的 l 趋于无穷,而且 R_j 的区间长度也都趋于零时,可用求极限的方法讨论有界的 NLAR(1)模型的平稳解存在问题. 这些内容涉及较多的概率测度弱收敛的知识,有待进一步研究,这里不再深入讨论了.

对一般的 NLAR 模型(4.2.27)或模型(4.2.29),求其不变概率分布只能给出如下的积分方程:

$$P(\boldsymbol{x}_1 \in A) = \int P(\boldsymbol{x}, A) P_0(\mathrm{d}\boldsymbol{x}) = \int P(\boldsymbol{x}, A) P(\mathrm{d}\boldsymbol{x}), \tag{4.4.36}$$

式中 $P(\boldsymbol{x}_t \in A)$ 为模型(4.2.29)的不变概率测度. 对于方程(4.4.36)的求解问题,一般只能用数值计算方法. 在极少的特殊情况下,也可能有特殊的求解方法,如在§4.2中对 LAR(1)模型(4.2.10)的不变概率分布的讨论即是如此. 下面再举一个非线性的例子.

例4.4.6 考虑如下 NLAR(1)模型

$$x_t = -\alpha |x_{t-1}| + \varepsilon_t, \tag{4.4.37}$$

式中$\{\varepsilon_t\}$为白噪声序列,满足(4.2.28)式. ε_t 具有连续且为正的密度 f_ε,并且$|\alpha|<1$. 由定理4.2.8知,模型(4.4.37)是几何遍历的. 另外,(4.4.37)式显然满足定理4.4.1的条件(4.4.7),从而只要$\{\varepsilon_t\}$为白噪声序列且满足(4.2.28)式,则(4.4.37)式有唯一平稳解. 这里不要求 ε_t 有处处正密度. 但是,如果 ε_t 有对称密度 f_ε 时,可求出(4.4.37)式的不变概率分布. 记模型(4.4.37)的不变概率分布的密度为 h,由(4.4.36)式有

$$\begin{aligned} h(y) &= \int h(x) f_\varepsilon(y + \alpha|x|)\mathrm{d}x \\ &= \int_0^\infty h(x) f_\varepsilon(y + \alpha x)\mathrm{d}x \\ &\quad + \int_{-\infty}^0 h(x) f_\varepsilon(y - \alpha x)\mathrm{d}x, \end{aligned} \tag{4.4.38}$$

由 f_ε 的对称性又有

$$h(-y) = \int_0^\infty h(x) f_\varepsilon(y - \alpha x)\mathrm{d}x$$

$$+\int_{-\infty}^{0} h(x)f_{\varepsilon}(y+\alpha x)\mathrm{d}x. \tag{4.4.39}$$

记 $\bar{h}(y)=h(y)+h(-y)$，由(4.4.38)和(4.4.39)可得

$$\bar{h}(y)=\int_{-\infty}^{\infty} \bar{h}(x)f_{\varepsilon}(y+\alpha x)\mathrm{d}x. \tag{4.4.40}$$

容易验证，(4.4.40)式中的 $\bar{h}$ 恰好是如下 LAR(1)模型的不变概率分布的密度

$$x_t=-\alpha x_{t-1}+\varepsilon_t. \tag{4.4.41}$$

对 f_ε 的某些特殊情况，我们还可进一步求出 h，例如当 ε_t 服从正态 $N(0,\sigma^2)$分布且$0<\alpha<1$时，由 § 4.2的讨论知，$\bar{h}$ 为正态 $N(0,\sigma^2/(1-\alpha^2))$分布的密度函数，即

$$\bar{h}(x)=\frac{1}{\sqrt{2\pi}}\sqrt{\frac{1-\alpha^2}{\sigma^2}}\exp\left\{-\frac{1}{2}\frac{1-\alpha^2}{\sigma^2}x^2\right\}.$$

再由(4.4.38)式，可得

$$\begin{aligned}
h(y)&=\int_0^{\infty} h(x)f_\varepsilon(y+\alpha x)\mathrm{d}x+\int_{-\infty}^{0} h(x)f_\varepsilon(y-\alpha x)\mathrm{d}x\\
&=\int_0^{\infty} h(x)f_\varepsilon(y+\alpha x)\mathrm{d}x+\int_0^{\infty} h(-x)f_\varepsilon(y+\alpha x)\mathrm{d}x\\
&=\int_0^{\infty} \bar{h}(x)f_\varepsilon(y+\alpha x)\mathrm{d}x\\
&=\sqrt{\frac{1-\alpha^2}{2\pi\sigma^2}}\cdot\frac{1}{\sqrt{2\pi}\sigma}\int_0^{\infty} e^{-\frac{1-\alpha^2}{2\sigma^2}x^2-\frac{(y+\alpha x)^2}{2\sigma^2}}\mathrm{d}x\\
&=\left\{\frac{1}{\sqrt{2\pi}}\int_0^{\infty} e^{-\frac{1}{2}\left(z+\frac{\alpha y}{\sigma}\right)^2}\mathrm{d}z\right\}\bar{h}(y)\\
&=\left\{\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-\frac{\alpha y}{\sigma}} e^{-\frac{t^2}{2}}\mathrm{d}t\right\}\bar{h}(y)\\
&=\Phi\left(-\frac{\alpha y}{\sigma}\right)\bar{h}(y),
\end{aligned} \tag{4.4.42}$$

式中 $\Phi(\cdot)$为 $N(0,1)$的分布函数. 再由 $\bar{h}(y)$的定义便得到

$$h(y)=\sqrt{\frac{(1-\alpha^2)}{2\pi\sigma^2}}\exp\left\{-\frac{(1-\alpha^2)x^2}{2\sigma^2}\right\}\Phi\left\{-\frac{\alpha x}{\sigma}\right\}.$$

最后，再讨论一个例子，它不满足定理4.2.8的压缩条件(4.2.31)式，但却是平稳遍历的.

例4.4.7 考虑如下的分段线性模型:

$$x_t = \begin{cases} x_{t-1} - 1 + \varepsilon_t, & \text{当 } x_{t-1} \geqslant 0; \\ x_{t-1} + 1 + \varepsilon_t, & \text{当 } x_{t-1} < 0, \end{cases} \tag{4.4.43}$$

式中$\{\varepsilon_t\}$为正态白噪声序列,$\varepsilon_t \sim N(0,\sigma^2)$,且$\varepsilon_t$与$x_{t-1}$相互独立. 显然,模型(4.4.43)中的变换$\varphi$为

$$\varphi(x) = \begin{cases} x - 1, & \text{当 } x \geqslant 0; \\ x + 1, & \text{当 } x < 0. \end{cases}$$

它不满足条件(4.2.31)式. 下面讨论模型(4.4.43)是否有平稳解的问题.

由(4.4.43)式有

$$\begin{aligned} E\left\{ |x_1| \,\middle|\, x_0 = x \right\} &= E\{ |\varphi(x_0) + \varepsilon_1| \,|\, x_0 = x\} \\ &= E|\varphi(x) + \varepsilon_1|. \end{aligned} \tag{4.4.44}$$

取准则函数$g(x) = |x|$. 为利用定理4.1.11,需要计算$E|a+\varepsilon_1|$的值:由$\varepsilon_1 \sim N(0,\sigma^2)$,有

$$E|a + \varepsilon_1| = E|a - \varepsilon_1|.$$

经计算,对$a>0$有

$$\begin{aligned} E|a + \varepsilon_1| &= \frac{1}{2}(E|a + \varepsilon_1| + E|a - \varepsilon_1|) \\ &= a + \int_{-\infty}^{-a} \left(\frac{|a + x| + |a - x|}{2} - a \right) \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} dx \\ &\quad + \int_a^{\infty} \left(\frac{|a + x| + |a - x|}{2} - a \right) \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} dx. \end{aligned} \tag{4.4.45}$$

由于

$$\begin{aligned} &\int_{-\infty}^{-a} \left(\frac{|a + x| + |a - x|}{2} - a \right) \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_0^{\infty} y e^{-(y-a)^2/2\sigma^2} dy \\ &\leqslant e^{-a^2/2\sigma^2} \frac{1}{\sqrt{2\pi}\sigma} \int_0^{\infty} y e^{-y^2/2\sigma^2} dy \leqslant \frac{\sigma}{2} e^{-a^2/2\sigma^2}, \end{aligned} \tag{4.4.46}$$

类似地有

$$\int_a^{\infty} \left(\frac{|a + x| + |a - x|}{2} - a \right) \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} dx \leqslant \frac{\sigma}{2} e^{-a^2/2\sigma^2}. \tag{4.4.47}$$

由(4.4.45)~(4.4.47)诸式,有

$$E|a+\varepsilon_1| \leqslant a + \sigma e^{-a^2/2\sigma^2}.$$

取 $a_0>0$足够大，使得当 $a>a_0$时，

$$E|a+\varepsilon_1| \leqslant a + \frac{1}{2}. \tag{4.4.48}$$

类似地，当 $a<-a_0$时，

$$E|a-\varepsilon_1| \leqslant |a| + \frac{1}{2}. \tag{4.4.49}$$

取 $A=[-a_0-1,a_0+1]$，由(4.4.45)、(4.4.48)和(4.4.49)诸式知

$$E\{|x_1|\,|x_0=x\} \leqslant |x| - \frac{1}{2}, \qquad \forall x \overline{\in} A,$$

$$E\{|x_1|\,|x_0=x\} \leqslant a_0 + \frac{1}{2} < \infty, \quad \forall x \in A.$$

于是，根据定理4.1.11知，由(4.3.43)式决定的马氏链是遍历的，从而有平稳解. 但不一定是几何遍历的. 验证非几何遍历是非常困难的，这里从略.

三、其他非线性时间序列模型

在前两小节中，主要讨论了某些可加噪声 NLAR 模型的平稳解. 在这一小节里，要讨论更一般的 NLARMA 模型. 对于这些模型，无论是平稳解的存在问题，还是解的遍历性问题，都不能使用 § 4.2中的马氏链方法. 所以这里不得不使用其他方法讨论这些模型的平稳解存在问题. 实际上，我们要介绍的方法，仍属于平稳 LAR 模型的后移方法，更确切地说，是使用随机变量求和的方法，并且将独立序列加权和推广到较一般的独立随机变量函数求和的情况.

(一) GARCH 模型的平稳遍历性

考虑 GARCH 模型(3.3.11)，即

$$\begin{cases} x_t=\varepsilon_t h_t^{1/2}; \\ h_t=\alpha_0+\alpha_1 x_{t-1}^2+\cdots+\alpha_p x_{t-p}^2+\beta_1 h_{t-1}+\cdots+\beta_q h_{t-q}, \end{cases} \tag{4.4.50}$$

式中$\{\varepsilon_t\}$为白噪声序列，满足(3.3.10)式，$p>0,q\geqslant 0,\alpha_0>0,\alpha_i\geqslant 0$ $(i=1,\cdots,p)$，$\beta_j\geqslant 0$ $(j=1,\cdots,q)$.

显然，当 $\beta_j=0$ $(j=1,\cdots,q)$时，模型(4.4.50)即为 ARCH 模型(4.3.14).

令 $y_t=x_t^2$，$\eta_t=\varepsilon_t^2$，$\boldsymbol{y}_t=(y_t,\ \cdots,\ y_{t-p+1},\ h_t,\ \cdots,\ h_{t-q+1})^\tau$，$\boldsymbol{b}_t=(\alpha_0\eta_t,0,\cdots,0,\alpha_0,0,\cdots,0)^\tau$，$A=EA_t$，其中

$$A_t=\begin{pmatrix}\alpha_1\eta_t & \cdots & \alpha_{p-1}\eta_t & \alpha_p\eta_t & \beta_1\eta_t & \cdots & \beta_{q-1}\eta_t & \beta_q\eta_t\\ 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0\\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0\\ \alpha_1 & \cdots & \alpha_{p-1} & \alpha_p & \beta_1 & \cdots & \beta_{q-1} & \beta_q\\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0\\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0\end{pmatrix}.$$

于是模型(4.4.50)可写成

$$y_t=\sum_{j=1}^{p}\alpha_j\eta_t y_{t-j}+\sum_{k=1}^{q}\beta_k\eta_t h_{t-k}+\alpha_0\eta_t,\tag{4.4.51}$$

其相应的一阶高维模型为

$$\boldsymbol{y}_t=A_t\boldsymbol{y}_{t-1}+\boldsymbol{b}_t.\tag{4.4.52}$$

引理 4.4.8　如果 $\alpha_1+\cdots+\alpha_p+\beta_1+\cdots+\beta_q<1$，则随机级数

$$\sum_{k=1}^{\infty}\left(\prod_{j=0}^{k-1}A_{t-j}\right)\boldsymbol{b}_{t-k}$$

以概率为 1 地收敛. 记

$$\boldsymbol{y}_t=\boldsymbol{b}_t+\sum_{k=1}^{\infty}\left(\prod_{j=0}^{k-1}A_{t-j}\right)\boldsymbol{b}_{t-k},\tag{4.4.53}$$

则$\{\boldsymbol{y}_t\}$是平稳随机序列，且满足随机差分方程(4.4.52)式.

证明　由 A_t 和 $\boldsymbol{b}_t$ 的定义易见，$\{A_t\}$和$\{\boldsymbol{b}_t\}$都是独立同分布的非负随机序列，且当 $k\neq j$ 时，A_{t-j}与 $\boldsymbol{b}_{t-k}$相互独立，再注意由(3.3.10)式知

$$E\varepsilon_t^2=E\eta_t=1,$$

于是有

$$E\left(\prod_{j=0}^{k-1}A_{t-j}\right)\boldsymbol{b}_{t-k}=\left(\prod_{j=0}^{k-1}EA_{t-j}\right)E\boldsymbol{b}_{t-k}=A^k\boldsymbol{b},\tag{4.4.54}$$

式中 $\boldsymbol{b}=(\alpha_0,0,\cdots,0,\alpha_0,0,\cdots,0)^{\tau}$，$A=EA_t$.

用矩阵论的方法，不难验证 $\rho(A)<1$，当且仅当 $\alpha_1+\cdots+\alpha_p+\beta_1+\cdots+\beta_q<1$. 故由引理的条件得

$$\sum_{k=1}^{\infty}A^k<\infty.\tag{4.4.55}$$

由(4.4.54)和(4.4.55)两式有

$$\sum_{k=1}^{\infty}E\left(\prod_{j=0}^{k-1}A_{t-j}\right)\boldsymbol{b}_{t-k}<\infty.$$

由于 A_t 和 $\boldsymbol{b}_t$ 都是非负的，从而由上式知，以概率为 1 地有

$$\sum_{k=1}^{\infty}\left(\prod_{j=0}^{k-1}A_{t-j}\right)\boldsymbol{b}_{t-k}<\infty.$$

显然，由(4.4.53)式定义的随机序列$\{\boldsymbol{y}_t\}$是平稳的. 进一步，通过直接验证可得

$$\begin{aligned}\boldsymbol{y}_t&=\boldsymbol{b}_t+A_t\left[\boldsymbol{b}_{t-1}+\sum_{k=2}^{\infty}\left(\prod_{j=0}^{k-1}A_{t-j}\right)\boldsymbol{b}_{t-k}\right]\\&=\boldsymbol{b}_t+A_t\left[\boldsymbol{b}_{t-1}+\sum_{l=1}^{\infty}\left(\prod_{j=0}^{l-1}A_{t-j}\right)\boldsymbol{b}_{t-1-l}\right]\\&=\boldsymbol{b}_t+A_t\boldsymbol{y}_{t-1}.\end{aligned}$$

引理得证. ▍

引理 4.4.9　如果$\{\boldsymbol{y}_t\}$是随机差分方程(4.4.52)式的平稳解，且 $E\boldsymbol{y}_0<\infty$，那么必有

$$\alpha_1+\cdots+\alpha_p+\beta_1+\cdots+\beta_q<1.$$

证明　由(4.4.52)式有

$$\begin{aligned}\boldsymbol{y}_0&=A_0\boldsymbol{y}_{-1}+\boldsymbol{b}_0\\&=\boldsymbol{b}_0+A_0\boldsymbol{b}_{-1}+A_0A_{-1}\boldsymbol{y}_{-2}\\&=\cdots\cdots\\&=\boldsymbol{b}_0+\sum_{k=1}^{n-1}\left(\prod_{j=0}^{k-1}A_{-j}\right)\boldsymbol{b}_{-k}+\left(\prod_{j=0}^{n-1}A_{-j}\right)\boldsymbol{y}_{-n}.\end{aligned}\tag{4.4.56}$$

由于 A_n、$\boldsymbol{b}_n$ 和 $\boldsymbol{y}_n$ 都是非负的，$\{A_t\}$和$\{\boldsymbol{b}_t\}$都是独立同分布的随机序列，且对 $k\neq j$，A_{n-j}与 $\boldsymbol{b}_{n-k}$相互独立，于是在引理的条件下，由(4.4.56)式知

$$E\boldsymbol{y}_0\geqslant\sum_{k=1}^{n-1}E\left(\prod_{j=0}^{k-1}A_{-j}\right)\boldsymbol{b}_{-k}=\sum_{k=1}^{n-1}A^k\boldsymbol{b}.$$

从而有

$$\sum_{k=1}^{\infty}A^k<\infty.$$

上式意味着

$$\lim_{n\to\infty}A^n=0.\tag{4.4.57}$$

由矩阵论的方法，容易验证(4.4.57)式蕴含了

$$\rho(A)<1,$$

从而得到 $\alpha_1+\cdots+\alpha_p+\beta_1+\cdots+\beta_q<1$. ▍

引理 4.4.10　设$\{\boldsymbol{y}_t\}$是随机差分方程(4.4.52)的平稳解，且 $E\boldsymbol{y}_0<$

∞. 如果还有一随机序列 $\{\boldsymbol{u}_t\}$，也是(4.4.52)式的平稳解，且 $E\boldsymbol{u}_0<\infty$，则必有

$$\boldsymbol{u}_t = \boldsymbol{y}_t, \quad \text{a.s.}.$$

证明 令 $\boldsymbol{v}_t=\boldsymbol{y}_t-\boldsymbol{u}_t$，则

$$\boldsymbol{v}_t = \boldsymbol{y}_t - \boldsymbol{u}_t = \left(\prod_{j=0}^{t-1} A_{t-j}\right)(\boldsymbol{y}_0 - \boldsymbol{u}_0).$$

为叙述方便，以下用 $(A)_{ij}$ 表示矩阵 A 的 (i,j) 元素，如果能证明对 $1\leqslant i\leqslant p+q$，存在常数 $c>0$ 和 $0<\lambda<1$，使得

$$E\left|\left[\left(\prod_{j=0}^{t-1} A_{t-j}\right)(\boldsymbol{y}_0 - \boldsymbol{u}_0)\right]_i\right| \leqslant c\lambda^t, \tag{4.4.58}$$

则由 Borel-Catelli 引理知

$$\lim_{t\to\infty} \boldsymbol{v}_t = \boldsymbol{0}, \quad \text{a.s.}.$$

从而得知 $\boldsymbol{y}_t=\boldsymbol{u}_t$ (a.s.).

由于证法类似，以下仅对 $i=1$ 证明(4.4.58)式成立. 由 A_t 的独立性，容易验证

$$E\left[\left(\prod_{j=0}^{t-1} A_{t-j}\right)_{1s}\right] = (A^t)_{1s}.$$

注意到 $\rho(A)<1$，存在常数 $\varepsilon>0$，满足 $\rho(A)+\varepsilon<1$ 及常数 $c=c(A,\varepsilon)$，使得

$$(A^t)_{1s} \leqslant c[\rho(A) + \varepsilon]^t.$$

因此

$$\begin{aligned} &E\left|\left[\left(\prod_{j=0}^{t-1} A_{t-j}\right)(\boldsymbol{y}_0 - \boldsymbol{u}_0)\right]_1\right| \\ &= E\left|\sum_{s=1}^{p}\left(\prod_{j=0}^{t-1} A_{t-j}\right)_{1s}(\boldsymbol{y}_0 - \boldsymbol{u}_0)_s\right| \\ &\leqslant \sum_{s=1}^{p} E\left[\left(\prod_{j=0}^{k-1} A_{t-j}\right)_{1s}\right] E|(\boldsymbol{y}_0 - \boldsymbol{u}_0)_s| \\ &\leqslant M[\rho(A) + \varepsilon]^t, \end{aligned}$$

式中常数

$$M = c[E|(\boldsymbol{y}_0 - \boldsymbol{u}_0)_1| + \cdots + E|(\boldsymbol{y}_0 - \boldsymbol{u}_0)_p|] < \infty.$$

引理得证. ▌

定理 4.4.11 GARCH 模型(4.4.51)有平稳解 $\{x_t\}$，且 $Ex_0^2<\infty$，当

且仅当 $\alpha_1+\cdots+\alpha_p+\beta_1+\cdots+\beta_q<1$. 此外,这个平稳解还是唯一的.

证明　由(4.4.50)、(4.4.51)和(4.4.52)三式可知,$\{x_t\}$是模型(4.4.50)的平稳解,当且仅当$\{\boldsymbol{y}_t\}$是模型(4.4.52)的平稳解. 为证明$\{x_t\}$的唯一性,只需证明(4.4.52)的平稳解$\{\boldsymbol{y}_t\}$的唯一性. 在(4.4.52)式中,从$\boldsymbol{y}_0=\boldsymbol{z}$出发,由(4.4.52)式反复迭代,得

$$\boldsymbol{y}_t^0=\boldsymbol{b}_t+\sum_{k=1}^{t-1}\left(\prod_{j=0}^{k-1}A_{t-j}\right)\boldsymbol{b}_{t-k}+\left(\prod_{j=0}^{t-1}A_{t-j}\right)\boldsymbol{z}.$$

由于 $\rho(A)<1$,类似于引理4.4.10的证明,可证

$$\left(\prod_{j=0}^{t-1}A_{t-j}\right)\boldsymbol{z}\to 0, \text{a.s.},\quad (\text{当 } t\to\infty),$$

从而 $\boldsymbol{y}_t^0$ 的分布 F_t 趋于(4.4.53)的平稳解的分布. 故模型(4.4.52)是唯一的,也即$\{\boldsymbol{y}_t\}$是唯一的. ▌

推论 4.4.12　ARCH 模型(4.3.14)有唯一平稳解,且 $Ex_t^2<\infty$,当且仅当 $\alpha_1+\cdots+\alpha_p<1$. 此外,平稳解$\{x_t\}$还是唯一的.

注意,在定理4.4.11和推论4.4.12中,并没有要求白噪声序列$\{\varepsilon_t\}$有密度函数. 比较推论4.4.12和定理4.3.6可看到,如果不涉及 ARCH 模型(4.3.14)的几何遍历性,则对$\{\varepsilon_t\}$"有处处为正的连续密度函数"的要求是可被减弱的. 这又一次可看到,研究 NLAR 模型平稳解时,使用 § 4.1 中的马氏链漂移准则的方法是有改进余地的.

(二) 双线性模型的平稳解

考虑双线性模型(3.4.4),即

$$x_t=\sum_{i=1}^{p}\varphi_i x_{t-i}+\sum_{j=0}^{q}\theta_j\varepsilon_{t-j}+\sum_{k=1}^{P}\sum_{l=1}^{Q}\beta_{kl}x_{t-k}\varepsilon_{t-l}, \qquad (4.4.59)$$

其中$\{\varepsilon_t\}$是独立同分布序列,满足(4.2.28)式. 记 $m=\max(P,p)$,

$$\boldsymbol{x}_t=(x_t,\cdots,x_{t-m+1})^{\tau},\quad \boldsymbol{\varepsilon}_t=(\varepsilon_t,\cdots,\varepsilon_{t-q})^{\tau},$$

$$\underset{m\times(q+1)}{C}=\begin{pmatrix}\theta_0 & \theta_1 & \cdots & \theta_q\\ 0 & 0 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 0\end{pmatrix},$$

$$\underset{m\times m}{A}=\begin{pmatrix}\varphi_1 & \varphi_2 & \cdots & \varphi_p & 0 & \cdots & 0 & 0\\ 1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0\\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & \cdots & 0 & 0 & \cdots & 1 & 0\end{pmatrix},$$

$$\underset{m\times m}{B_l}=\begin{pmatrix}\beta_{1l} & \beta_{2l} & \cdots & \beta_{pl} & 0 & \cdots & 0\\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0\end{pmatrix},$$

则模型(4.4.59)可写成如下的矩阵形式:

$$\boldsymbol{x}_t = C\boldsymbol{\varepsilon}_t + A\boldsymbol{x}_{t-1} + \sum_{l=1}^{Q} B_l \boldsymbol{x}_{t-1}\varepsilon_{t-l}. \tag{4.4.60}$$

为讨论模型(4.4.60)解的平稳性,先叙述如下引理:

引理 4.4.13 对如上定义的矩阵 A,存在一个 $m\times m$ 非奇异矩阵 R,使得

$$\|R^{-1}AR\|_\infty = \rho_0,$$

其中 $\rho_0=\rho(A)$. 对 $m\times m$ 矩阵 $G=(g_{ij})$,矩阵范数$\|\cdot\|_\infty$定义为:

$$\|G\|_\infty = \max_{1\leqslant i\leqslant m}\sum_{j=1}^{m}|g_{ij}|.$$

定理 4.4.14 设 A、B_l 和 R 如引理4.4.13所述. 令 $\delta_l=\|R^{-1}B_lR\|_\infty$,$1\leqslant l\leqslant Q$,如果

$$E\log\left(\rho_0 + \sum_{l=1}^{Q}\delta_l|\varepsilon_l|\right) < 0, \tag{4.4.61}$$

则随机级数

$$\sum_{r=1}^{\infty}\left[\prod_{u=0}^{r-1}\left(A+\sum_{l=1}^{Q}B_l\varepsilon_{t-u-l}\right)\right]C\boldsymbol{\varepsilon}_{t-r}$$

以概率为1地收敛. 进一步,记

$$\boldsymbol{x}_t = C\boldsymbol{\varepsilon}_t + \sum_{r=1}^{\infty}\left[\prod_{u=0}^{r-1}\left(A+\sum_{l=1}^{Q}B_l\varepsilon_{t-u-l}\right)\right]C\boldsymbol{\varepsilon}_{t-r}, \tag{4.4.62}$$

则$\{\boldsymbol{x}_t\}$是(4.4.60)式的唯一平稳解.

证明 令 $z_t=\gamma^\tau\boldsymbol{x}_t$,这里 γ 是任意的 m 维非零实数向量,再令

$$w_{rt} = \gamma^\tau\left[\prod_{u=0}^{r-1}\left(A+\sum_{l=1}^{Q}B_l\varepsilon_{t-u-l}\right)\right]C\boldsymbol{\varepsilon}_{t-r},$$

如果能证明随机级数

$$\sum_{r=1}^{\infty}w_{rt} \tag{4.4.63}$$

以概率为1地收敛,则定理的结论正确. 注意到

$$|w_{rt}|\leqslant M\|R\|_\infty\left[\prod_{u=0}^{r-1}\left\|R^{-1}\left(A+\sum_{l=1}^{Q}B_l\varepsilon_{t-u-l}\right)R\right\|_\infty\right]$$

$$\times \|R^{-1}\|_\infty \|\boldsymbol{\varepsilon}_{t-r}\|_\infty, \tag{4.4.64}$$

其中对任意 m 维向量 $\boldsymbol{u}=(u_1,\cdots,u_m)^{\tau}$ 定义向量范数 $\|\cdot\|_\infty$ 为 $\|\boldsymbol{u}\|_\infty=\max\{|u_1|,\cdots,|u_m|\}$，$M$ 是一个不依赖 r 的常数. 由(4.4.64)式和大数定律，有

$$\begin{aligned} r^{-1}\log|w_{rt}| \leqslant{} & r^{-1}\log M \\ & + r^{-1}\sum_{u=0}^{r-1}\log\left(\rho_0+\sum_{l=1}^{Q}\delta_l|\varepsilon_{t-u-l}|\right) \\ & + r^{-1}\sum_{u=0}^{q}\log|\varepsilon_{t-r-u}| \\ & \to E\log\left(\rho_0+\sum_{l=1}^{Q}\delta_l|\varepsilon_t|\right), \text{ a.s. （当 } r\to\infty \text{ 时）}. \end{aligned}$$

故由(4.4.61)式知，对每一个 t，有

$$\limsup_{r\to\infty}|w_{rt}|^{1/r}<1, \text{ a.s.}.$$

上式蕴含了(4.4.63)式的收敛. 通过直接验证可知，由(4.4.62)式确定的 $\{\boldsymbol{x}_t\}$ 是(4.4.60)式的解，而且是平稳的. 又由(4.4.60)式知，$\{x_t\}$ 是模型(4.4.59)的平稳解. 类似于定理4.4.11的证法，可证模型(4.4.59)的平稳解是唯一的. ▌

下面讨论模型(4.4.59)的几个特例.

例4.4.15　考虑如下简单的一阶双线性模型

$$x_t=\alpha x_{t-1}\varepsilon_{t-1}+\varepsilon_t, \tag{4.4.65}$$

其中 $\{\varepsilon_t\}$ 是独立同分布序列，满足(4.2.28)式. 对模型(4.4.65)，存在唯一平稳解的充分条件(4.4.61)式成为

$$E\log(|\alpha||\varepsilon_1|)<0. \tag{4.4.66}$$

如果进一步假定 $\{\varepsilon_t\}$ 是白噪声序列，满足(3.2.2)式，且 $|\alpha\sigma|<1$，那么用 Jenson 不等式和 Shwartz 不等式，有

$$\begin{aligned} E\log|\alpha\varepsilon_1| &\leqslant \log E|\alpha\varepsilon_1| \leqslant \log(\alpha^2E\varepsilon_1^2)^{1/2} \\ &= \log|\alpha\sigma|<0, \end{aligned}$$

即条件(4.4.66)成立. 从而模型(4.4.65)有唯一的平稳解，且平稳解可表为

$$x_t=\varepsilon_t+\sum_{r=1}^{\infty}\alpha^r\left(\prod_{u=0}^{r-1}\varepsilon_{t-1-u}\right)\varepsilon_{t-r}. \tag{4.4.67}$$

比较(4.4.67)式和(1.3.3)式可看出，(4.4.67)式中的 x_t 是一种特殊的

Volterra 展式.

虽然(4.4.67)式给出了模型(4.4.65)的平稳解的明显表达式，但是利用(4.4.67)式并不能给出 x_t 的概率分布与 ε_t 的概率分布的简单关系. 目前尚未有成熟的方法求出模型(4.4.65)的不变分布.

例4.4.16 考虑如下双线性模型

$$x_t = \alpha x_{t-1} + \varepsilon_t + \beta x_{t-1}\varepsilon_{t-1}, \qquad (4.4.68)$$

其中$\{\varepsilon_t\}$为白噪声序列，满足(4.2.28)式. 此时(4.4.61)式为

$$E\log(|\alpha| + |\beta||\varepsilon_1|) < 0. \qquad (4.4.69)$$

在(4.4.69)式下，模型(4.4.68)有唯一的平稳解. 如果假定

$$|\alpha| + |\beta| \cdot |\sigma| < 1, \qquad (4.4.70)$$

用类似于例4.4.15中的方法可以验证条件(4.4.69)式成立，从而模型(4.4.68)有平稳解，而且其平稳解可表为

$$x_t = \varepsilon_t + \sum_{r=1}^{\infty}\left[\prod_{u=0}^{r-1}(\alpha + \beta\varepsilon_{t-1-u})\right]\varepsilon_{t-r}. \qquad (4.4.71)$$

例4.4.17 考虑如下双线性模型

$$x_t = \sum_{i=1}^{p}\varphi_i x_{t-i} + \varepsilon_t + \sum_{j=1}^{p}\beta_j x_{t-j}\varepsilon_{t-1}, \qquad (4.4.72)$$

其中$\{\varepsilon_t\}$是独立同分布序列，满足(4.2.28)式. 令

$$\underset{p\times p}{A} = \begin{pmatrix} \varphi_1 & \cdots & \varphi_{p-1} & \varphi_p \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix},$$

$$\underset{p\times p}{B} = \begin{pmatrix} \beta_1 & \cdots & \beta_p \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}$$

$$\underset{p\times 1}{C} = (1, 0, \cdots, 0)^{\tau}, \qquad \boldsymbol{x}_t = (x_t, \cdots, x_{t-p+1})^{\tau},$$

则模型(4.4.72)的一阶高维形式为

$$\boldsymbol{x}_t = C\varepsilon_{t-1} + A\boldsymbol{x}_{t-1} + B\boldsymbol{x}_{t-1}\varepsilon_{t-1}. \qquad (4.4.73)$$

根据定理4.4.14，模型(4.4.72)存在唯一平稳解的充分条件为

$$E\log(\rho_0 + \delta_1|\varepsilon_1|) < 0, \qquad (4.4.74)$$

其中 $\rho_0 = \|R^{-1}AR\|_* = \rho(A)$，非奇异矩阵 R 按引理4.4.13定义，

$$\delta_0 = \|R^{-1}BR\|_*.$$

另外，如果假定$\{\varepsilon_t\}$为白噪声序例，满足(3.2.2)式，且

$$\rho(A\otimes A+\sigma^2 B\otimes B)<1,$$

其中"$\otimes$"为 Kronecker 积，用类似于定理4.4.14及其引理的证明方法，可以证明模型(4.4.73)有平稳解，而且平稳解是唯一的.

对于另一类双线性模型(3.4.5)，即

$$x_t=\theta_0+\sum_{i=1}^{q}\theta_i\varepsilon_{t-i}+\sum_{j=1}^{P}\sum_{k=1}^{Q}\beta_{jk}\varepsilon_{t-i}\varepsilon_{t-j}+\varepsilon_t,$$

式中x_t已经是$\{\varepsilon_t\}$的二次函数的形式. 在$\{\varepsilon_t\}$为白噪声序列的假定下，$\{x_t\}$总是平稳的. 但是，根据ε_t的概率分布来确定模型(3.4.5)的不变概率分布的问题仍是十分困难的.

(三) 双重线性模型的平稳解

考虑如下一阶双重线性模型

$$\begin{cases}x_t=(\alpha+\beta_t)x_{t-1}+\varepsilon_t;\\ \beta_t=\eta_t+\theta\eta_{t-1},\end{cases}\tag{4.4.75}$$

其中假定：

(i) $\{\varepsilon_t\}$为白噪声序列，满足(3.2.2)式，且$E\varepsilon_t^4<\infty$；

(ii) $\{\eta_t\}$为白噪声序列，$E\eta_t=0, E\eta_t^2=\sigma_\eta^2, E\eta_t^3=0, E\eta_t^4=\mu_4<\infty$，且$\{\eta_t\}$与$\{\varepsilon_t\}$相互独立.

由(4.4.75)式可得

$$x_t=\varepsilon_t+\sum_{i=1}^{n}\left[\prod_{j=0}^{i-1}(\alpha+\beta_{t-j})\right]\varepsilon_{t-i}+\prod_{j=0}^{n}(\alpha+\beta_{t-j})x_{t-n-j}.\tag{4.4.76}$$

令

$$S_n=E\left[\prod_{j=0}^{n}(\alpha+\beta_j)^2\right],\tag{4.4.77}$$

由概率论知识知，如果

$$\sum_{n=1}^{\infty}S_n<\infty,\tag{4.4.78}$$

则随机级数

$$\sum_{i=1}^{\infty}\left[\prod_{j=0}^{i-1}(\alpha+\beta_{t-j})\right]\varepsilon_{t-i}$$

是依均方收敛的，从而(4.4.76)式的x_t可表为

$$x_t = \varepsilon_t + \sum_{i=1}^{\infty}\left[\prod_{j=0}^{i-1}(\alpha + \beta_{t-j})\right]\varepsilon_{t-i}. \tag{4.4.79}$$

再由(4.4.79)式知$\{x_t\}$是平稳的.

下面讨论使(4.4.78)式成立的条件. 为此令

$$T_n = E\eta_n^2\left[\prod_{j=0}^{n}(\alpha + \beta_j)^2\right], \quad W_n = E\eta_n\left[\prod_{j=0}^{n}(\alpha + \beta_j)^2\right].$$

由(4.4.77)式,并注意到β_n的定义及条件(i)和(ii),便有

$$\begin{aligned} S_n &= E(\alpha + \beta_n)^2\prod_{j=0}^{n-1}(\alpha + \beta_j)^2 \\ &= E(\alpha^2 + 2\alpha\beta_n + \beta_n^2)\prod_{j=0}^{n-1}(\alpha + \beta_j)^2 \\ &= \alpha^2 S_{n-1} + 2\alpha E\beta_n\prod_{j=0}^{n-1}(\alpha + \beta_j)^2 \\ &\quad + E\beta_n^2\prod_{j=0}^{n-1}(\alpha + \beta_j)^2 \\ &= (\alpha^2 + \sigma_\eta^2)S_{n-1} + \theta^2 T_{n-1} + 2\alpha\theta W_{n-1}. \end{aligned} \tag{4.4.80}$$

用类似的方法,同理可证

$$T_n = (\alpha^2\sigma_\eta^2 + \mu_4)S_{n-1} + \theta^2\sigma_\eta^2 T_{n-1} + 2\alpha\theta\sigma_\eta^2 W_{n-1}, \tag{4.4.81}$$

$$W_n = 2\alpha\sigma_\eta^2 S_{n-1} + 2\theta\sigma_\eta^2 W_{n-1}. \tag{4.4.82}$$

记 $\boldsymbol{y}_n = (S_n, T_n, W_n)^\tau$,

$$A = \begin{pmatrix} \alpha^2 + \sigma_\eta^2 & \theta^2 & 2\alpha\theta \\ \alpha^2\sigma_\eta^2 + \mu_4 & \theta^2\sigma_\eta^2 & 2\alpha\theta\sigma_\eta^2 \\ 2\alpha\sigma_\eta^2 & 0 & 2\theta\sigma_\eta^2 \end{pmatrix}.$$

则由(4.4.80)～(4.4.82)三式可得到如下迭代关系：

$$\boldsymbol{y}_n = A\boldsymbol{y}_{n-1}. \tag{4.4.83}$$

用类似引理4.4.9的证法可证明,只要(4.4.83)中的矩阵A满足$\rho(A) < 1$,则必有(4.4.78)式成立,从而(4.4.75)式有平稳解.

用上述方法研究更一般的双重线性模型的平稳解的存在问题,会遇到非常复杂的各种高阶矩的条件,这里就不再深入讨论了,读者可参阅文献[123].

§ 4.5 非线性时间序列模型的高阶矩

在时间序列的统计分析中，模型的平稳性、遍历性和高阶矩的存在性是非常重要的. 因为在时间序列模型的估计理论中，无论是参数估计方法，还是非参数估计方法，或者是半参数估计方法，都需要模型有平稳遍历性，且存在适当阶的矩. 本章前4节主要讨论非线性时间序列模型的平稳性和遍历性. 本节将讨论非线性时间序列模型的矩的性质，特别是可加噪声的 NLAR 模型、函数型随机条件方差的 NLAR 模型，以及 GARCH 模型的高阶矩的存在条件. 为解决 NLAR 模型高阶矩的存在问题，一般状态马氏链的方法仍然是不可缺少的工具. 确切地说，我们要利用如下引理来检验 NLAR 模型高阶矩的存在性，引理的证明见文献[48].

引理4.5.1 设$\{\boldsymbol{x}_t\}$是取值于$(R^p,\mathscr{B}_p)$的遍历的马氏链，π 是$\{\boldsymbol{x}_t\}$的不变概率分布. 如果存在非负可测的 p 元函数 g 和某一个小集 C，使得

(i) $0<g(\boldsymbol{x})<\infty$， $\forall\ \boldsymbol{x}\in R^p$；

(ii) $\sup\limits_{\boldsymbol{x}\in C}E\{g(\boldsymbol{x}_t)\mid\boldsymbol{x}_{t-1}=\boldsymbol{x}\}<\infty$； (4.5.1)

(iii) 存在正常数 $\delta<1$，使得

$$E\{g(\boldsymbol{x}_t)\mid\boldsymbol{x}_{t-1}=\boldsymbol{x}\}\leqslant(1-\delta)g(\boldsymbol{x}),\forall\ \boldsymbol{x}\overline{\in}C,\tag{4.5.2}$$

那么

$$\int_{R^p}g(\boldsymbol{x})\pi(\mathrm{d}\boldsymbol{x})<\infty,\tag{4.5.3}$$

而且对某个正常数 $\rho<1$，当 $t\to\infty$ 时，有

$$\int_{R^p}\pi(\mathrm{d}\boldsymbol{x})\sup_{|h|\leqslant g}\left|E\{h(\boldsymbol{x}_t)\mid\boldsymbol{x}_0=\boldsymbol{x}\}-\int_{R^p}\pi(\mathrm{d}\boldsymbol{y})h(\boldsymbol{y})\right| = O(\rho^t).\tag{4.5.4}$$

在下面的论述中，我们还要使用矩阵的 Kronecker 积和拉直运算. 记

$$A^{\otimes r}=\underbrace{A\otimes A\otimes\cdots\otimes A}_{r次}.$$

并记 $\boldsymbol{a}_i(i=1,2,\cdots,m)$为矩阵 A 的列向量，则矩阵拉直运算记为

$$\mathrm{Vec}(A)=(a_1^{\tau},\cdots,a_m^{\tau})^{\tau},$$

并有如下的关系式：

$$\rho(A^{\otimes r})=\rho^r(A)$$

和

$$\mathrm{Vec}(ABC)=(C^{\tau}\otimes A)\mathrm{Vec}(B).\tag{4.5.5}$$

定理4.5.2 设时间序列$\{x_t\}$服从NLAR模型(4.2.27),并满足推论4.2.9的条件(i)和(ii).如果$\rho(A)<1$,其中矩阵A由(4.2.36)式定义,r为某个正整数,且$E|\varepsilon_t|^r<\infty$,则$E|x_t|^r<\infty$.

证明 记$\boldsymbol{x}_t=(x_t,\cdots,x_{t-p+1})^\tau$,则$\{\boldsymbol{x}_t\}$服从模型(4.2.27)相应的一阶$p$维模型(4.2.29).在定理的条件下,由推论4.2.9知$\{\boldsymbol{x}_t\}$是几何遍历的.定义

$$W_r(t)=\underbrace{\mathrm{Vec}(\cdots\mathrm{Vec}(\mathrm{Vec}}_{r\text{次}}(\boldsymbol{x}_t\boldsymbol{x}_t^\tau)\otimes\boldsymbol{x}_t^\tau)\cdots\otimes\boldsymbol{x}_t^\tau),$$

为证明$E|x_t|^r<\infty$,只需证明$E\|W_r(t)\|_v<\infty$,其中$\|\cdot\|_v$为一个适当选取的向量范数.不失一般性,仅考虑$r=2$的情形.

对给定的$\varepsilon>0$,且满足$\rho^2(A)+\varepsilon<1$,存在由某个向量范数$\|\cdot\|_v$诱导出的矩阵范数$\|\cdot\|_{mv}$,使得

$$\rho^2(A)\leqslant\|A^{\otimes 2}\|_{mv}\leqslant\rho^2(A)+\varepsilon.\tag{4.5.6}$$

(见文献[67]).

另外,由有限维向量范数的等价性,存在常数$0<\delta_1<\delta_2<\infty$,使得

$$\delta_1\|\mathrm{Vec}(\boldsymbol{x}\boldsymbol{x}^\tau)\|\leqslant\|\mathrm{Vec}(\boldsymbol{x}\boldsymbol{x}^\tau)\|_v\leqslant\delta_2\|\mathrm{Vec}(\boldsymbol{x}\boldsymbol{x}^\tau)\|.\tag{4.5.7}$$

利用(4.2.36)式的表达式及其中出现的矩阵A、向量$\boldsymbol{u}$和函数$\boldsymbol{H}(\boldsymbol{x})$等记号,记

$$\begin{aligned}D(\boldsymbol{x})=&\|\mathrm{Vec}(A\boldsymbol{x}(\boldsymbol{H}(\boldsymbol{x}))^\tau)\|_v\\&+\|\mathrm{Vec}(\boldsymbol{H}(\boldsymbol{x})\boldsymbol{x}^\tau A^\tau)\|_v\\&+\|\mathrm{Vec}(\boldsymbol{H}(\boldsymbol{x})\boldsymbol{H}(\boldsymbol{x})^\tau)\|_v\\&+E\{\|\mathrm{Vec}(A\boldsymbol{x}\boldsymbol{u}^\tau\varepsilon_1)\|_v\\&+\|\mathrm{Vec}(\boldsymbol{H}(\boldsymbol{x})\boldsymbol{u}^\tau\varepsilon_1)\|_v\\&+\|\mathrm{Vec}(\varepsilon_1\boldsymbol{u}\boldsymbol{x}^\tau A^\tau)\|_v\\&+\|\mathrm{Vec}(\varepsilon_1\boldsymbol{u}(\boldsymbol{H}(\boldsymbol{x}))^\tau)\|_v\},\end{aligned}$$

取非负函数$g(\boldsymbol{x})=1+\|\mathrm{Vec}(\boldsymbol{x}\boldsymbol{x}^\tau)\|_v$,再利用(4.2.36)式,经直接验算可得到

$$\begin{aligned}E\{g(\boldsymbol{x}_{t+1})|\boldsymbol{x}_t=\boldsymbol{x}\}&=E\{g(\boldsymbol{x}_1)|\boldsymbol{x}_0=\boldsymbol{x}\}\\&=1+E\{\|\mathrm{Vec}((A\boldsymbol{x}_0+\boldsymbol{H}(\boldsymbol{x}_0)+\varepsilon_1\boldsymbol{u})\\&\qquad\times(A\boldsymbol{x}_0+\boldsymbol{H}(\boldsymbol{x}_0)+\varepsilon_1\boldsymbol{u})^\tau)\|_v|\boldsymbol{x}_0=\boldsymbol{x}\}\\&\leqslant\|\mathrm{Vec}(A\boldsymbol{x}\boldsymbol{x}^\tau A^\tau)\|_v+D(\boldsymbol{x})+c\\&\leqslant\|A^{\otimes 2}\|_{mv}\|\mathrm{Vec}(\boldsymbol{x}\boldsymbol{x}^\tau)\|_v+D(\boldsymbol{x})+c\end{aligned}$$

$$\leqslant \{\rho^2(A)+\varepsilon\}\|\mathrm{Vec}(\boldsymbol{x}\boldsymbol{x}^\tau)\|_v + D(\boldsymbol{x}) + c, \qquad (4.5.8)$$

其中 c 为某个正常数. 由推论4.2.9的条件(i)和(4.5.7)式,容易验证

$$\frac{D(\boldsymbol{x})}{\|\mathrm{Vec}(\boldsymbol{x}\boldsymbol{x}^\tau)\|} \to 0,\ 当\ \|\mathrm{Vec}(\boldsymbol{x}\boldsymbol{x}^\tau)\| \to \infty\ 时.$$

这样,当$\|\mathrm{Vec}(\boldsymbol{x}\boldsymbol{x}^\tau)\|$足够大时,(4.5.8)式右端可被$(1-\delta)g(\boldsymbol{x})$控制,其中正常数 δ 满足$0<\rho(A)+\varepsilon<\delta<1$. 取

$$C = \{\boldsymbol{x}:\|\mathrm{Vec}(\boldsymbol{x}\boldsymbol{x}^\tau)\| \leqslant K\},$$

其中 K 是一个足够大的正数,使得当$\|\mathrm{Vec}(\boldsymbol{x}\boldsymbol{x}^\tau)\|\geqslant K$ 时,有

$$E\{g(\boldsymbol{x}_{t+1})|\boldsymbol{x}_t=\boldsymbol{x}\} \leqslant (1-\delta)g(\boldsymbol{x}).$$

根据引理4.2.5,C 是小集,而且存在 $c_1>0$,使得

$$E\{g(\boldsymbol{x}_{t+1})|\boldsymbol{x}_t=\boldsymbol{x}\} \leqslant c_1 < \infty,\quad \forall\ \boldsymbol{x}\in C.$$

故由引理4.5.1知

$$\int_{R^2} g(\boldsymbol{x})\pi(\mathrm{d}\boldsymbol{x}) < \infty.$$

上式蕴含了 $E|x_t|^2<\infty$. ▌

定理4.5.3　设时间序列$\{x_t\}$满足 NLAR 模型(4.2.27),且定理4.2.14的条件成立,则 $E|\varepsilon_t|^r<\infty$蕴含了 $E|x_t|^r<\infty$,其中 r 为正整数.

定理4.5.3的证明与定理4.5.2类似,这里从略.

定理4.5.4　设时间序列$\{x_t\}$满足 NLAR 模型(4.2.60),且定理4.2.18的条件成立. 如果 $\rho(A)<1$,且 $E|\varepsilon_t|^r<\infty$,其中 r 为某个正整数,则 $E|x_t|^r<\infty$.

用类似于定理4.2.18和定理4.5.2的证明方法可证明定理4.5.4,这里从略.

现在讨论 §4.3中所述的某些 NLAR 模型的高阶矩性质.

例4.5.5　考虑例4.3.3的指数型 NLAR 模型(3.3.6),假定其白噪声序列$\{\varepsilon_t\}$还满足 $E|\varepsilon_t|^r<\infty$,其中 $r>0$为整数,且自回归系数 $\boldsymbol{\alpha}=(\varphi_{01},\cdots,\varphi_{0p})^\tau$满足(4.2.33)式,记

$$A=\begin{pmatrix}\varphi_{01} & \cdots & \varphi_{0p-1} & \varphi_{0p}\\ 1 & \cdots & 0 & 0\\ \vdots & \ddots & \vdots & \vdots\\ 0 & \cdots & 1 & 0\end{pmatrix},$$

如果 $\rho(A)<1$,其中 $r>0$为整数,则由定理4.5.2知 $E|x_t|^r<\infty$.

例4.5.6　考虑半参数模型(4.3.7),即

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + h(x_{t-1}, \cdots, x_{t-p}) + \varepsilon_t,$$

其中自回归系数 $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_p)^\tau$ 满足(4.2.33)式,函数 h 满足(4.3.8)和(4.3.9)式. 如果$\{\varepsilon_t\}$满足定理4.5.2的条件,且矩阵

$$A = \begin{bmatrix} \alpha_1 & \cdots & \alpha_{p-1} & \alpha_p \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix}$$

的谱半径 $\rho(A) < 1$,则 $E|x_t|^r < \infty$.

例 4.5.7　考虑门限自回归模型(4.3.25),如果自回归系数满足(4.3.26)式,且 ε_t 还具有有穷的 r 阶矩,则由定理4.5.3知,模型(4.3.25)也有有穷的 r 阶矩.

例 4.5.8　考虑广义 LAR 模型(4.3.21),即

$$x_t = \varphi(\theta_1 x_{t-1} + \cdots + \theta_p x_{t-p}) + \varepsilon_t,$$

如果函数 φ 满足(4.3.22)式,则由定理4.5.3知 $E|\varepsilon_t|^r < \infty$ 蕴含了

$$E|x_t|^r < \infty.$$

如果函数 φ 满足(4.3.23)式,则由定理4.5.2知 $E|\varepsilon_t|^r < \infty$ 蕴含了

$$E|x_t|^r < \infty.$$

例 4.5.9　考虑 ANLAR 模型(4.3.4),即

$$x_t = c + f_1(x_{t-1}) + \cdots + f_p(x_{t-p}) + \varepsilon_t,$$

且满足例4.3.2的条件,如果 $E|\varepsilon_t|^r < \infty$,则由定理4.5.3知 $E|x_t|^r < \infty$.

例 4.5.10　考虑 β-ARCH 模型(4.3.11),即

$$x_t = \varepsilon_t(\alpha_0 + \alpha_1 x_{t-1}^{2\beta} + \cdots + \alpha_p x_{t-p}^{2\beta})^{1/2},$$

其中$0 \leqslant \beta < 1$. 如果例4.3.5中的条件被满足,而且 $E|\varepsilon_t|^r < \infty$,则由定理4.5.3知 $E|x_t|^r < \infty$.

例 4.5.11　考虑随机条件方差门限自回归模型(4.3.37),即

$$x_t = \sum_{j=1}^{l} \left\{ \varphi_{j0} + \sum_{k=1}^{p} \varphi_{jk} x_{t-k} + \varepsilon_{jt} \right\} I(x_{t-d} \in R_j),$$

其中记号和条件如例4.3.15所假定. 仿照定理4.2.14和定理4.5.2的证明可证明(4.3.26)和 $E|\varepsilon_{jt}|^r < \infty$ $(j = 1, \cdots, l)$蕴含了 $E|x_t|^r < \infty$.

对上述诸例所讨论的与 NLAR 模型相应的随机条件方差模型,利用定理4.5.4可讨论其高阶矩的存在性.

还需指出,如同§4.2和§4.3中所提到的那样,使用一般状态马氏链

方法时，对白噪声 ε_t 的分布的限制较强. 例如要求 ε_t 的分布有处处连续的正密度. 要放宽这种限制，还有待进一步研究.

对于其他非线性时间序列模型，例如 GARCH 模型、双线性模型等，虽然不能直接使用引理4.5.1来讨论其矩的性质，但是却可以利用其平稳解的明显表达式来直接讨论它们的矩的性质. 下面先来讨论 GARCH 模型的高阶矩. 仍沿用 § 4.4中的记号，并令

$$L^s = \{x : \|x\|_s = E^{1/s}|x|^s < \infty, s > 0\},$$

其中 x 是随机变量.

定理4.5.12　设时间序列$\{x_t\}$服从 GARCH 模型(4.4.50). 假定

$$E|\varepsilon_t|^{4(s-1)} < \infty,$$

其中 $s\geqslant2$为整数，且 $\alpha_1+\cdots+\alpha_p+\beta_1+\cdots+\beta_q<1$. 记

$$\sum\nolimits_m = E(A_t^{\otimes m}).$$

如果 $\rho(\sum_s) < \infty$，则$|x_t|^2\in L^s$.

证明　不失一般性，考虑 $s=2$的情形，$s>2$的证明完全类似. 记

$$S_n(t)=\boldsymbol{b}_t+\sum_{k=1}^{n}\Big(\prod_{j=0}^{k-1}A_{t-j}\Big)\boldsymbol{b}_{t-k}, \tag{4.5.9}$$

$$W_{n+m,n}(t)=S_{n+m}(t)-S_n(t),\ m\geqslant1,$$

$$\Gamma_{n+m,n}(t)=\mathrm{Vec}(W_{n+m,n}(t)W_{n+m,n}^{\tau}(t)).$$

则容易验证

$$W_{n+m,n}(t)=A_tW_{n-1+m,n-1}(t-1). \tag{4.5.10}$$

并由上式及矩阵 Kronecker 积和拉直运算的性质，有

$$\begin{aligned}\Gamma_{n+m,n}(t)&=\mathrm{Vec}(A_tW_{n-1+m,n-1}(t-1)W_{n-1+m,n-1}^{\tau}(t-1)A_t^{\tau})\\&=(A_t^{\otimes2})\mathrm{Vec}(W_{n-1+m,n-1}(t-1)W_{n-1+m,n-1}^{\tau}(t-1))\\&=(A_t^{\otimes2})\Gamma_{n-1+m,n-1}(t-1).\end{aligned} \tag{4.5.11}$$

显然，$\Gamma_{n-1+m,n-1}(t-1)$是 $\mathscr{F}_{t-1}=\sigma(A_s,b_s,s\leqslant t-1)$可测的. 注意到$\{A_t\}$是独立同分布的，由(4.5.11)式可得到

$$\begin{aligned}E\Gamma_{n+m,n}(t)&=E(A_t^{\otimes2})E\Gamma_{n-1+m,n-1}(t-1)\\&=(E(A_t^{\otimes2}))^2E\Gamma_{n-2+m,n-2}(t-2)\\&=\cdots\cdots\\&=(E(A_t^{\otimes2}))^nE\Gamma_{m,0}(t-n).\end{aligned} \tag{4.5.12}$$

但是

$$\begin{aligned}
\Gamma_{m,0}(t-n) &= \mathrm{Vec}(W_{m,0}(t-n)W_{m,0}^{\tau}(t-n)), \\
W_{m,0}(t-n) &= S_{m,0}(t-n) - S_0(t-n) \\
&= \sum_{k=1}^{m}\left(\prod_{j=0}^{k-1}A_{t-j}\right)\boldsymbol{b}_{t-n-k} \\
&\triangleq (w_1,\cdots,w_{p+q})^{\tau},
\end{aligned}$$

其中

$$w_i = \sum_{k=1}^{m}\alpha_0\left[\left(\prod_{j=0}^{k-1}A_{t-n-j}\right)_{i1}\eta_{t-n-k} + \left(\prod_{j=0}^{k-1}A_{t-n-j}\right)_{ip+1}\right].$$

于是,注意到

$$\begin{aligned}
&\left(\left(\prod_{j=0}^{k-1}A_{t-n-j}\right)_{i1}\right)^2 \\
&= \left(\left(\prod_{j=0}^{k-1}A_{t-n-j}\right)\otimes\left(\prod_{j=0}^{k-1}A_{t-n-j}\right)\right)_{i1,i1} \\
&= \left(\prod_{j=0}^{k-1}(A_{t-n-j}\otimes A_{t-n-j})\right)_{i1,i1},
\end{aligned}$$

这里记号$(\cdot)_{i1,i1}$表示原来矩阵中$(i,1)$元素在 Kronecker 积中的位置. 由 Hölder 不等式和 Minkowski 不等式,有

$$\begin{aligned}
(Ew_i^2)^{1/2} &= \left\{E\left\{\sum_{k=1}^{m}\alpha_0\left[\left(\prod_{j=0}^{k-1}A_{t-n-j}\right)_{i1}\eta_{t-n-k}\right.\right.\right. \\
&\qquad \left.\left.\left.+\left(\prod_{j=0}^{k-1}A_{t-n-j}\right)_{ip+1}\right]\right\}^2\right\}^{1/2} \\
&\leqslant \sum_{k=1}^{m}\alpha_0\left\{E\left[\left(\prod_{j=0}^{k-1}A_{t-n-j}\right)_{i1}\eta_{t-n-k}\right.\right. \\
&\qquad \left.\left.+\left(\prod_{j=0}^{k-1}A_{t-n-j}\right)_{ip+1}\right]^2\right\}^{1/2} \\
&\leqslant \sum_{k=1}^{m}\alpha_0\left\{\left\{E\left[\left(\prod_{j=0}^{k-1}A_{t-n-j}\right)_{i1}\eta_{t-n-k}\right]^2\right\}^{1/2}\right. \\
&\qquad \left.+\left\{E\left[\left(\prod_{j=0}^{k-1}A_{t-n-j}\right)_{ip+1}\right]^2\right\}^{1/2}\right\} \\
&\leqslant \sum_{k=1}^{m}\alpha_0\left\{\left\{E\left[\left(\prod_{j=0}^{k-1}A_{t-n-j}\right)_{i1}\right]^2\right\}^{1/2}\right.
\end{aligned}$$

$$\times \left\{ E\eta_{t-n-k}^2 \right\}^{1/2} + \left\{ E\left[\left(\prod_{j=0}^{k-1} A_{t-n-j} \right)_{i1} \right]^2 \right\}^{1/2} \Bigg\}$$

$$\leqslant c \sum_{k=1}^{m} \{ \{ [(E(A_t^{\otimes 2}))^k]_{i1,i1} \}^{1/2}$$

$$+ \{ [(E(A_t^{\otimes 2}))^k]_{ip+1,ip+1} \}^{1/2} \}. \tag{4.5.13}$$

由于 $\lambda = \rho(\Sigma_2) < 1$，故存在常数 $c > 0$，使得

$$(\Sigma_2^k)_{i1,i1} \leqslant c \{ \rho(\Sigma_2) \}^k.$$

于是，由(4.5.13)式有

$$\{ Ew_i^2 \}^{1/2} \leqslant c \sum_{k=1}^{m} \lambda^{k/2} < \infty. \tag{4.5.14}$$

令

$$\mu_{2m} = \max_{1 \leqslant i \leqslant (p+q)^2} \{ E(\Gamma_{m,0}(t-n))_i \},$$

则由(4.5.14)式有

$$E[\Gamma_{n+m,n}(t)]_1 \leqslant c\lambda^n. \tag{4.5.15}$$

令 $\boldsymbol{\delta} = (1, 0, \cdots, 0)^\tau$，类似于(4.5.12)式的证明，可得到

$$\begin{aligned}
[\boldsymbol{\delta}^\tau W_{n+m,n}(t)]^2 &= \boldsymbol{\delta}^\tau W_{n+m,n}(t) W_{n+m,n}^\tau(t) \boldsymbol{\delta} \\
&= \mathrm{Vec}(\boldsymbol{\delta}^\tau W_{n+m,n}(t) W_{n+m,n}^\tau(t) \boldsymbol{\delta}) \\
&= (\boldsymbol{\delta}^{\otimes 2}) \mathrm{Vec}(W_{n+m,n}(t) W_{n+m,n}^\tau(t)) \\
&= [\Gamma_{n+m,n}(t)]_1.
\end{aligned}$$

这样，由(4.5.15)式得

$$\begin{aligned}
\| \boldsymbol{\delta}^\tau W_{n+m,n}(t) \| &= \{ E[\boldsymbol{\delta}^\tau W_{n+m,n}(t)]^2 \}^{1/2} = \{ E[\Gamma_{n+m,n}(t)]_1 \}^{1/2} \\
&\leqslant c\lambda^{n/2} \to 0, \text{ 当 } n \to \infty, m \to \infty \text{ 时},
\end{aligned}$$

这就意味着
$$x_{t,n}^2 = \boldsymbol{\delta}^\tau \left[\boldsymbol{b}_t + \sum_{k=1}^{n} \left(\prod_{j=0}^{k-1} A_{t-j} \right) \boldsymbol{b}_{t-k} \right]$$

是 L^2 中的 Cauchy 基本列. 故 $x_t^2 \in L^2$. ▌

推论 4.5.13　设时间序列 $\{x_t\}$ 服从 ARCH 模型(4.3.14)，假定 $E|\varepsilon_t|^{4(s-1)} < \infty$，其中 $s \geqslant 2$ 是整数，且 $\alpha_1 + \cdots + \alpha_p < 1$. 记

$$\Sigma_s = E(A_t^{\otimes s}),$$

其中
$$A_t = \begin{pmatrix} \alpha_1 \eta_t & \cdots & \alpha_{p-1} \eta_t & \alpha_p \eta_t \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix}.$$

如果 $\rho(\Sigma_s)<1$，则 $x_t^2\in L^s$.

现在讨论双线性模型(4.4.59)的矩的性质. 对一般模型(4.4.59)，其高阶矩存在的条件较为复杂，这里仅讨论 $q=Q=1$的情形，即

$$x_t = \sum_{i=1}^{p}\varphi_i x_{t-i} + \sum_{k=1}^{P}\beta_{k1}x_{t-k}\varepsilon_{t-1} + \varepsilon_t. \qquad (4.5.16)$$

沿用(4.4.60)式的记号，(4.5.15)式可写为

$$\boldsymbol{x}_t = A\boldsymbol{x}_{t-1} + B_1\boldsymbol{x}_{t-1}\varepsilon_{t-1} + \varepsilon_t.$$

记 $\gamma_4\triangleq E\varepsilon_t^4$，

$$\begin{aligned}\Lambda\triangleq\ & A^{\otimes 4} + \sigma^2(A\otimes B_1)^{\otimes 2} + \sigma^2(B_1\otimes A)^{\otimes 2}\\ & + \sigma^2(B_1^{\otimes 2}\otimes A^{\otimes 2}) + \sigma^2(B_1\otimes A)\otimes(A\otimes B_1)\\ & + \sigma^2(A\otimes B_1)\otimes(B_1\otimes A) + \sigma^2(A^{\otimes 2}\otimes B_1^{\otimes 2})\\ & + \gamma_4 B_1^{\otimes 4}.\end{aligned}$$

用类似于定理4.5.12的证明方法可证明如下结论：

定理 4.5.14　设时间序列$\{x_t\}$满足模型(4.5.16)，其中白噪声序列$\{\varepsilon_t\}$还满足

$$E\varepsilon_t = E\varepsilon_t^3 = 0,\ E\varepsilon_t^2 = \sigma^2,\ E\varepsilon_t^4 = \gamma_4 < \infty, \qquad (4.5.17)$$

$$\lambda = \rho(A\otimes A + \sigma^2 B_1\otimes B_1) < 1. \qquad (4.5.18)$$

如果 $\rho(\Lambda)<1$，则 $E|x_t|^4<\infty$.

例 4.5.15　考虑双线性模型(4.4.65)，即

$$x_t = \alpha x_{t-1}\varepsilon_{t-1} + \varepsilon_t,$$

其中白噪声序列$\{\varepsilon_t\}$满足(4.5.17)式，且$|\alpha\sigma|<1$. 如果 $\alpha^4E\varepsilon_t^4<\infty$，则由定理4.5.14知 $Ex_t^4<\infty$.

例 4.5.16　考虑双线性模型(4.4.68)，即

$$x_t = \alpha x_{t-1} + \beta x_{t-1}\varepsilon_{t-1} + \varepsilon_t,$$

其中白噪声序列$\{\varepsilon_t\}$满足(4.5.17)，且 $\alpha^2+\beta^2\sigma^2<1$. 如果

$$\alpha^4 + 6\sigma^2\alpha^2\beta^2 + \gamma_4\beta^4 < 1, \qquad (4.5.19)$$

则由定理4.5.14知 $Ex_t^4<\infty$.

§ 4.6　时间序列的两种可逆性

本章前几节所讨论的 NLAR 模型的平稳解是将 x_t 表成$\{\varepsilon_t,\varepsilon_{t-1},\cdots\}$的某种函数，例如熟知的 LAR(1)模型

$$x_t = \alpha x_{t-1} + \varepsilon_t. \qquad (4.6.1)$$

当 $|\alpha|<1$时,模型(4.6.1)有平稳解

$$x_t = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j},$$

即将 x_t 表示成$\{\varepsilon_t,\varepsilon_{t-1},\cdots\}$的指数加权和的函数. 本节将要讨论的时间序列的第一种可逆性,是指能将 ε_t 表示成$\{x_t,x_{t-1},\cdots\}$的某种函数. 在 LARMA 模型的研究中已有熟知的结果(见文献[125]). 如在 § 1.2中,为了保证 LARMA 模型(1.2.4)的 ε_t 能表示成(1.2.11)式,需要一定的条件——即可逆性条件(1.2.6)式. 在此顺便指出:如果用$\{x_t,x_{t-1},\cdots\}$的某种函数的极限表示 ε_t,也称为可逆的,并不一定非要用无穷级数(1.2.11)式来表示 ε_t. 这样,可逆性条件(1.2.6)式还可以放宽为

$$1-\theta_1 z-\cdots-\theta_q z^q \neq 0, |z|<1. \tag{4.6.2}$$

这一结论要涉及较多的泛函分析的知识(参见文献[131]). 由于形如(1.2.11)式的表达式在统计分析时极为方便,所以在研究 LARMA 模型的文献中多使用(1.2.6)式的可逆条件.

下面讨论 NLARMA 模型的可逆性问题. 考虑 NLARMA 模型(3.4.2),即

$$x_t=\varphi(x_{t-1},\cdots,x_{t-p},\varepsilon_{t-1},\cdots,\varepsilon_{t-q})+\varepsilon_t, \tag{4.6.3}$$

其中$\{\varepsilon_t\}$为白噪声序列,满足(3.2.2)式,φ 是 R^{p+q}到 R^1的可测函数.

定义 4.6.1　设 NLARMA 模型(4.6.3)有唯一平稳解. 若存在$\{x_t,x_{t-1},\cdots\}$的函数,记为 $\psi\{x_t,x_{t-1},\cdots\}$,使得(4.6.3)式中的 ε_t 能表示成

$$\varepsilon_t=\psi\{x_t,x_{t-1},\cdots\}, \tag{4.6.4}$$

则称 NLARMA 模型(4.6.3)是可逆的,也称$\{x_t\}$为可逆的.

不像 LARMA 模型(1.2.4)那样,只要 MA 的系数 $\boldsymbol{\theta}=(\theta_1,\cdots,\theta_q)^\tau$ 满足(1.2.6)式或(4.6.2)式,该模型就是可逆的. 对于 NLARMA 模型(4.6.3),一般地说,模型类型不同,其可逆性条件也不同. 因此以下将通过几个例子来简略地讨论 NLARMA 模型的可逆性.

例 4.6.2　当(4.6.3)式为可加噪声 NLAR 模型(4.2.27)时,即

$$x_t=\varphi(x_{t-1},\cdots,x_{t-p})+\varepsilon_t,$$

显然

$$\varepsilon_t=x_t-\varphi(x_{t-1},\cdots,x_{t-p}) \triangleq \psi\{x_t,\cdots,x_{t-p}\}. \tag{4.6.5}$$

(4.6.5)式表明,可加噪声 NLAR 模型是无条件可逆的.

例 4.6.3　函数型随机条件方差 NLAR 模型(4.2.24),即

$$x_t = \varphi(x_{t-1},\cdots,x_{t-p}) + \varepsilon_t S(x_{t-1},\cdots,x_{t-p}). \qquad (4.6.6)$$

由于 S 是正值函数，故

$$\varepsilon_t = \frac{x_t - \varphi(x_{t-1},\cdots,x_{t-p})}{S(x_{t-1},\cdots,x_{t-p})}. \qquad (4.6.7)$$

(4.6.7)式表明了模型(4.6.6)也是可逆的.

例4.6.4 考虑双线性模型(4.4.65)，即

$$x_t = \varepsilon_t - \alpha x_{t-1}\varepsilon_{t-1}, \qquad (4.6.8)$$

其中$\{\varepsilon_t\}$是白噪声序列，满足(3.2.2)式. 于是有

$$\varepsilon_t = x_t + \alpha x_{t-1}\varepsilon_{t-1}. \qquad (4.6.9)$$

由(4.6.8)和(4.6.9)的对应关系，即将 x_t 与 ε_t 互换后，仿照(4.4.67)式，可得如下关于 ε_t 的表达式：

$$\varepsilon_t = x_t + \sum_{k=1}^{n-1}\alpha^k\left(\prod_{i=1}^{k}x_{t-i}\right)x_{t-k} + \alpha^n\cdot\left(\prod_{k=1}^{n}x_{t-k}\right)\varepsilon_{t-n}. \qquad (4.6.10)$$

一般说来，讨论(4.6.10)式求和的收敛问题比(4.4.67)式困难得多，原因在于 $x_{t-1},x_{t-2},\cdots$的相关结构很复杂. 但对一些特殊情况，例如 ε_t 服从$(-\delta,\delta)$上的均匀分布，其中$\frac{2}{4}<\delta<\frac{3}{4}$，且(4.6.9)式中的 α 满足$|\alpha|<(1-\delta)<1$，可以验证，(4.6.10)式是依概率收敛的，从而模型(4.6.8)是可逆的. 事实上，在上述条件下，易见(4.4.66)式成立，从而模型(4.6.8)有唯一的平稳解. 由(4.4.67)式和$|\varepsilon_t|\leqslant\delta$，有

$$\begin{aligned}|x_t| &\leqslant |\varepsilon_t| + \sum_{k=1}^{\infty}|\alpha|^k\left(\prod_{i=1}^{k}|\varepsilon_{t-i}|\right)|\varepsilon_{t-k}| \\ &\leqslant \delta + \sum_{k=1}^{\infty}|\alpha|^k\delta^{k+1} \\ &= \delta + \frac{|\alpha|\delta^2}{(1-|\alpha|)} < 1.\end{aligned}$$

将上式代入(4.6.10)式中可知，(4.6.10)式最后一项依概率趋于零，从而以下无穷级数

$$\varepsilon_t = x_t + \sum_{k=1}^{\infty}\alpha^k\left(\prod_{i=1}^{k}x_{t-i}\right)x_{t-k}$$

依概率收敛.

例4.6.5　考虑如下有理分式模型

$$x_t = \frac{x_{t-1}\varepsilon_t}{x_{t-1}^2 + \varepsilon_t^2}, \tag{4.6.11}$$

式中$\{\varepsilon_t\}$为正态白噪声，$\varepsilon_t \sim N(0,\sigma^2)$. 由文献[106]中 Tong 的结果知，模型(4.6.11)是几何遍历的，从而有唯一的平稳解. 但是当给定$\{x_t, x_{t-1}, \cdots\}$时，由(4.6.11)式知，ε_t 应满足如下的一元二次方程式

$$x_t\varepsilon_t^2 - x_{t-1}\varepsilon_t + x_{t-1}^2 x_t = 0. \tag{4.6.12}$$

注意到 ε_t 与 x_{t-1}独立，且 $\varepsilon_t \sim N(0,\sigma^2)$，由于(4.6.12)式中的系数是自由参数，从而 ε_t 不是总能被(x_t, x_{t-1})所唯一确定. 故模型(4.6.11)不是可逆的.

时间序列模型的可逆性在统计分析中是很有用的. 这在 LARMA 模型的统计分析中已得到充分体现(参见文献[16]). 在非线性时间序列模型中也是如此. 当模型具有可逆性时，不可观测的白噪声序列$\{\varepsilon_t\}$可以用可观测的时间序列$\{x_t\}$进行估计，从而能估计一步预报的条件方差和无条件方差. 进一步还可以利用$\{\varepsilon_t\}$的独立同分布假定以及 ε_t 与$\{x_s, s<t\}$独立的假定，来构造各种检验统计量，以解决各种检验问题，例如在第6章中的线性性检验问题.

现在讨论另一种意义的可逆性——时间可逆性(Time Reversibilily).

定义4.6.6　平稳序列$\{x_t\}$称为时间可逆的，如果对于正整数 m 和任何时刻 $t_1<t_2<\cdots<t_m$ 以及任意时刻 s，$(x_{t_1}, \cdots, x_{t_m})$与$(x_{s-t_1}, \cdots, x_{s-t_m})$有相同的分布.

注意，如果$\{x_t\}$是平稳序列，则对任何固定的 s，$\{x_{s-t}\}$也是平稳序列. 故在定义4.6.6中的 s 可以是任意的时刻. 特别地，可取 $s=0$. 当被讨论的平稳序列限定在 $t\geqslant 1$的时刻内，即$\{x_t, t\geqslant 1\}$是平稳序列，在定义4.6.6中应取 $s>t_m$.

直观地说，定义4.6.6中的可逆性是指时间反向后的序列，与原序列的概率结构相同.

如果平稳序列$\{x_t\}$是零均值正态序列，则由定义4.6.6易知，它一定是时间可逆的. 事实上，对于任意的正整数 m 和任意时刻 $t_1<t_2<\cdots<t_m$，$(x_{t_1}, \cdots, x_{t_m})^\tau$ 的分布只与它的协方差阵

$$\Gamma = (\gamma_{t_i - t_j})_{i,j=1,\cdots,m} \tag{4.6.13}$$

有关. 根据平稳性知，对任意时刻 s，$(x_{s-t_1}, \cdots, x_{s-t_m})^\tau$ 的协方差阵也是(4.6.13)式的 Γ. 所以它与$(x_{t_1}, \cdots, x_{t_m})^\tau$ 有相同的分布.

从实际背景来看，一个时间序列往往是客观现象随时间发展的观测序列，它反映了某种物理、化学或经济等现象的变化过程. 一般说来，将时间反向后的被观测序列并不一定再保持原序列的概率结构. 所以时间可逆性并不具有广泛的实际意义. 当然，如果一个平稳序列具有时间可逆性，这也是一种很有意义的结构特征. 在 LARMA 模型的拟合研究中，这种特征曾被用来提高参数估计的精度(参见文献[22]、[133]).

在非线性时间序列中，不仅存在时间可逆的平稳序列，也存在着时间不可逆的平稳序列. 但是，对于一般的 NLAR 模型和 NLARMA 模型，很难判断其相应的平稳序列是否为时间可逆的. 故而代替讨论与之有关的纯随机序列与单向随机序列. 需要指出的是，在非线性时间序列中，还会出现时间反向后为非随机序列的现象. 我们先看以下例子.

例4.6.7 考虑如下的 LAR(1)模型：

$$x_t = \frac{1}{2}x_{t-1} + \frac{1}{2}\varepsilon_t, \tag{4.6.14}$$

式中$\{\varepsilon_t\}$为独立同分布序列，$P(\varepsilon_t=0)=P(\varepsilon_t=1)=\frac{1}{2}$. 由§4.2中关于模型(4.2.42)的讨论知模型(4.6.14)具有平稳解，而且平稳分布为[0,1]上的均匀分布. 根据定义4.6.1，模型(4.6.14)是可逆的，而且 ε_t 可表为

$$\varepsilon_t = 2x_t - x_{t-1}. \tag{4.6.15}$$

为了说明模型(4.6.14)不是时间可逆的，首先计算在给定 x_{t-1}时，x_t 的条件分布. 由(4.6.14)式，易得

$$P\left(x_t = \frac{1}{2}a \,\middle|\, x_{t-1} = a\right) = \frac{1}{2}, \quad \forall\, a \in [0,1],$$

$$P\left(x_t = \frac{1}{2}a + \frac{1}{2} \,\middle|\, x_{t-1} = a\right) = \frac{1}{2}, \quad \forall\, a \in [0,1].$$

此两个式子表明了在给定 $x_{t-1}=a(0\leqslant a\leqslant 1)$时，$x_t$ 的条件分布不退化. 但是若将时间反向，即考虑在给定 x_t 时，x_{t-1}的条件分布，由(4.6.14)式并注意到 x_t 服从[0,1]上的均匀分布，则有

当 $x_t=a<\frac{1}{2}$时，必有 $\varepsilon_t=0$，从而 $x_{t-1}=2a$； (4.6.16)

当 $x_t=a\geqslant\frac{1}{2}$时，必有 $\varepsilon_t=1$，从而 $x_{t-1}=2a-1$. (4.6.17)

记

$$f_1(x) = 2x\,\mathrm{mod}(1) = \begin{cases} 2x, & \text{当 } 0 \leqslant x \leqslant \frac{1}{2}; \\ 2x-1, & \text{当 } \frac{1}{2} \leqslant x \leqslant 1. \end{cases} \tag{4.6.18}$$

联合(4.6.16)～(4.6.18)三个式子，得

$$x_{t-1} = f_1(x_t) = 2x_t \mathrm{mod}(1). \tag{4.6.19}$$

(4.6.19)式表明了当时间反向后，x_{t-1}被x_t所唯一确定. 因此在给定x_t时，x_{t-1}的条件分布是退化的. 可见此时$\{x_t\}$不是时间可逆的.

例4.6.8　考虑如下的双线性模型：

$$x_t = \frac{1}{2}(1 - \varepsilon_t + \varepsilon_t x_{t-1}), \tag{4.6.20}$$

式中$\{\varepsilon_t\}$为独立同分布序列，且$P(\varepsilon_t=1)=P(\varepsilon_t=-1)=\frac{1}{2}$. 通过直接验证可知，[0,1]上的均匀分布为模型(4.6.20)的不变分布. 由例4.4.15可知，模型(4.6.20)有唯一的平稳解. 当将时间反向后如何呢?用例4.6.7的推导方法可以得到

$$x_{t-1} = f_2(x_t), \tag{4.6.21}$$

其中

$$f_2(x) = \begin{cases} 2x, & \text{当 } 0 \leqslant x < \frac{1}{2}; \\ 2 - 2x, & \text{当 } \frac{1}{2} \leqslant x \leqslant 1. \end{cases} \tag{4.6.22}$$

例4.6.9　考虑如下的非线性模型：

$$x_t = \frac{1}{2} + \frac{1}{2}\varepsilon_t(1 - x_{t-1})^{1/2}, \tag{4.6.23}$$

式中$\{\varepsilon_t\}$为独立同分布序列，且

$$P(\varepsilon_t = 1) = P(\varepsilon_t = -1) = \frac{1}{2}.$$

通过直接验证可知，Bata 分布 $B(0.5,0.5)$是模型(4.6.23)的不变概率分布，即其不变概率分布具有密度

$$f_B(x) = \frac{1}{\sqrt{\pi(1-x)x}}, \quad 0 < x < 1. \tag{4.6.24}$$

从而(4.6.23)式有平稳解. 于是根据例4.6.3知，(4.6.23)是可逆的，而且

$$\varepsilon_t = \frac{2x_{t-1} - 1}{\sqrt{1 - x_{t-1}}}.$$

但是，当将时间反向后，利用例4.6.8的讨论方法可得到

$$x_{t-1} = f_3(x_t), \tag{4.6.25}$$

其中

$$f_3(x)=4x(1-x),\quad x\in(0,1).$$

上述三个平稳序列在时间反向后都表现为确定性的序列，即它们分别满足某确定性的迭代关系式，或称为离散动态系统. 这又一次显示出非线性平稳序列与动态系统的联系. 为了进一步揭示它们之间的联系，先来分析一下如何将确定性序列反向后变成一个随机序列.

考虑满足(4.6.19)式的确定性序列：

$$\begin{cases}y_t=2y_t\text{mod}(1);\\ y_0\in[0,1].\end{cases}\tag{4.6.26}$$

现在要问：在给定 y_t 的值时，能否确定 y_{t-1} 的值. 由(4.6.19)式的函数 $f_1(x)$ 的图象(见图2.2.2)可看到，f_1 的逆不是唯一的，即任给 $y_t=a\in(0,1)$，y_{t-1} 有两个可能值，即 $y_{t-1}=\frac{1}{2}a$ 或者 $\frac{1}{2}a+\frac{1}{2}$. 如果我们用一个 Bernoulli 试验决定 $y_{t-1}=\frac{a}{2}$ 还是 $y_{t-1}=\frac{a}{2}+\frac{1}{2}$，并各以概率 $\frac{1}{2}$ 取每个值，于是可将 y_{t-1} 写成

$$y_{t-1}=\frac{1}{2}y_t+\frac{1}{2}\varepsilon_t,\tag{4.6.27}$$

其中 $\{\varepsilon_t\}$ 是独立同分布序列，$P(\varepsilon_t=0)=P(\varepsilon_t=1)=\frac{1}{2}$. 于是(4.6.27)与(4.6.14)为同一模型. 这就从另一方面在 LAR(1)模型与确定性动态系统(4.6.26)之间建立了联系.

在上述讨论中，如果取 $P(\varepsilon_t=0)=p$，$P(\varepsilon_t=1)=1-p$，则由(4.6.27)式仍可用 Bernoulli 试验决定 y_{t-1} 的取值，而 y_t 与 y_{t-1} 仍保持 $y_t=2y_{t-1}\text{mod}(1)$ 的关系. 于是，自然会想到，为什么取 $p=\frac{1}{2}$ 呢?换言之，$p=\frac{1}{2}$ 有什么特殊性质呢?对不同的 p 值. 只要 $0<p<1$，模型(4.6.27)就有不同的平稳分布. 而在动态系统(4.6.26)的多个不变分布中，只有[0,1]上的均匀分布是特殊的遍历分布. 这个不变分布与模型(4.6.27)中 $p=\frac{1}{2}$ 的情形相对应. 关于动态系统的不变分布和遍历性，我们将在第8章中讨论，这里暂不仔细论述了.

以上的讨论不难推广到例4.6.8和例4.6.9. 更一般地说，若 y_t 满足一个确定性迭代系统，即

$$\begin{cases}y_t=f(y_{t-1}),\quad t\geqslant 1;\\ y_0\in R^1,\end{cases}\tag{4.6.28}$$

当函数 f 不是单值时，则 y_{t-1} 不能被 y_t 所唯一确定. 特别是，当 y_{t-1} 具有有限个不同的取值时，总可以用一个取有限个值的离散随机变量 ε_t，以其不同的取值来确定对应的 y_{t-1} 的取值. 于是便得到了关于 $\{y_t\}$ 的反向的自回归模型. 至于 ε_t 取各种可能值的概率，则可根据系统(4.6.28)的遍历的不变分布而定. 由于(4.6.28)式的遍历的不变分布很难给出来，所以很难确定上述 ε_t 的分布，除非对某些简单的模型——例如对例4.6.7、例4.6.8和例4.6.9等有可能给出 ε_t 的合理的概率值.

综上所述，我们可以利用时间序列的正向和反向的随机性，将平稳时间序列分为如下两类：

定义 4.6.10　如果平稳时间序列 $\{x_t\}$ 不满足下列两条件中的任何一条：

(i) 存在函数 ξ，使得

$$x_k = \xi(x_{k-1}, x_{k-2}, \cdots), \ \forall\ k; \tag{4.6.29}$$

(ii) 存在函数 ζ，使得

$$x_k = \zeta(x_{k+1}, x_{k+2}, \cdots), \ \forall\ k. \tag{4.6.30}$$

则称 $\{x_t\}$ 为*纯随机平稳序列*.

定义 4.6.11　如果平稳时间序列 $\{x_t\}$ 满足定义4.6.10中的条件(i)或(ii)，则称 $\{x_t\}$ 为*单向随机平稳序列*.

注意，若一个平稳序列 $\{x_t\}$ 满足定义4.6.10中的条件(i)或条件(ii)，不妨设满足条件(ii)，则由平稳性知

$$x_t = \zeta(x_{t+1}, x_{t+2}, \cdots), \ \forall\ t \geqslant 1. \tag{4.6.31}$$

这表明了当将时间反向时，平稳序列 $\{x_t\}$ 满足(4.6.31)式的确定性动态系统. 关于同时满足(4.6.29)和(4.6.30)式的序列，将在第 8 章中讨论.

第 5 章

非线性时间序列的预报

时间序列分析的一个重要内容是预报.时间序列的预报有着广泛的应用背景,无论是线性时间序列,还是非线性时间序列,都是如此.但是对这两类时间序列,解决其预报问题的方法和困难程度却有很大差别.在本章中,将对比线性平稳时间序列预报与非线性平稳时间序列预报的相同与相异之处,并介绍几种重要的非线性平稳时间序列模型的预报方法.本章主要讨论平稳时间序列的预报问题,为了简便,若无特别说明,常略去"平稳"二字.

§5.1 时间序列预报的概述

所谓时间序列的预报,实质上就是根据时间序列的历史观测值 $x_n, x_{n-1}, \cdots$,对未来时刻 $n+k(k>0)$ 的取值进行估计.以 $\{x_n\}$ 表示一个时间序列,不妨假定 $Ex_n=0, Ex_n^2<\infty$.由§1.4的讨论可知,在给定历史值 $\{x_n, x_{n-1}, \cdots\}$ 的条件下,使用不同的方法给出将来时刻 $n+k$ 的取值的估计,将导致不同的预报方法.而估计 $x_{n+k}(k>0)$ 可以有不同的准则.从概率论角度看,最完备的预报应当是给出在 $\{x_n, x_{n-1}, \cdots\}$ 的条件下 x_{n+k} 的条件分布.但由于条件分布既不容易被估计,又不便于使用,所以较少被讨论.代之用条件分布的数字特征——条件期望作为 x_{n+k} 的估计,称之为**条件期望预报方法**,这是一种重要的预报方法.在非线性时间序列的预报中尤其如此.

如§1.4所讨论的那样,在给定 $\{x_n, x_{n-1}, \cdots\}$ 的条件下,利用条件期望的性质,x_{n+1} 的条件期望可表示为

$$\begin{aligned}\hat{x}_n(1) &= E\{x_{n+1} \mid x_n, x_{n-1}, \cdots\} \\ &= \varphi(x_n, x_{n-1}, \cdots).\end{aligned} \tag{5.1.1}$$

记

$$e_{n+1} = x_{n+1} - \hat{x}_n(1), \quad n \geqslant 0, \tag{5.1.2}$$

称为时间序列 $\{x_n\}$ 的**新息序列**.在§1.4中曾指出,新息序列 $\{e_n\}$ 是平稳

鞅差序列，根据§1.4中的性质1.4.1，有

$$E\{e_{n+1}|x_n,x_{n-1},\cdots\}=0,\tag{5.1.3}$$

$$Ee_{n+1}^2=\inf_{\xi}E\{x_{n+1}-\xi(x_n,x_{n+1},\cdots)\}^2,\tag{5.1.4}$$

上式中的 ξ 取遍所有的可测函数. 如§3.1中所述，按照(5.1.1)式中的预报函数 φ，可对时间序列进行分类. 若 φ 只依赖于有限个 $x_n,x_{n-1},\cdots,x_{n-p+1}$，则由(5.1.1)和(5.1.2)两式得如下模型：

$$x_t=\varphi(x_{t-1},\cdots,x_{t-p})+e_t.\tag{5.1.5}$$

这是一个有穷阶自回归模型. 一般说来，时间序列 $\{x_n\}$ 可能是如下的无穷阶自回归模型的解，即

$$x_t=\varphi(x_{t-1},x_{t-2},\cdots)+e_t.\tag{5.1.6}$$

在此必须再次强调：正如§1.4中指出的那样，无论是在(5.1.5)式中还是在(5.1.6)式中，对 $\{e_t\}$ 的最自然的假定只能是平稳鞅差序列. 但是，在第3章和第4章中又曾指出，为使模型有明确的定义，必须假定 $\{e_t\}$ 为白噪声序列，否则，就会使(5.1.5)式包括了一大类等价模型. 其原因在于：用条件期望预报 x_{n+1} 时，并未完全限定 x_{n+1} 的条件分布，从而在模型(5.1.5)中，尽管具有相同的条件期望预报函数，但其新息序列 $\{e_t\}$ 可以有许多不同的更精确的概率结构. 为此，当 $\{e_t\}$ 为平稳鞅差序列时，应当称(5.1.5)式为**预报 AR(p)模型**. 当 $\{e_t\}$ 为白噪声序列时，则称(5.1.5)式为**结构 AR(p)模型**，按模型(3.2.1)的表示，即

$$x_t=\varphi(x_{t-1},x_{t-2},\cdots,x_{t-p})+\varepsilon_t.\tag{5.1.7}$$

以条件期望作为预报方法，可以保证其预报误差的方差有极小性质，即(5.1.4)式成立. 但是，便于计算的一种实用的预报方法是用历史资料 $\{x_n,x_{n-1},\cdots\}$ 的线性最小方差方法来预报 x_{n+k}. 这一方法被称为**投影预报方法**. 当给定 $\{x_n,x_{n-1},\cdots\}$ 时，x_{n+k} 的投影预报按如下准则确定，即

$$\inf_{c_1,c_2,\cdots}E\Big\{x_{n+k}-\sum_{j=0}^{\infty}c_jx_{n-j}\Big\}^2\quad(k=1,2,\cdots).\tag{5.1.8}$$

相应的预报值记为

$$\hat{x}_n^*(k)=\hat{E}\{x_{n+k}|x_n,x_{n-1},\cdots\}.\tag{5.1.9}$$

在本节中，我们只讨论以上两种预报方法，借助这两种预报方法来进一步讨论线性时间序列与非线性时间序列的差异. 在§1.4中曾指出，当 $\{x_t\}$ 为(1.2.1)式的线性序列时，条件期望预报与投影预报是一致的，即 $\hat{x}_n(1)=\hat{x}_n^*(1)$. 当 $\{x_t\}$ 为非线性时间序列时，§1.4中的例1.4.2表明了 $\hat{x}_n(1)\neq\hat{x}_n^*(1)$ 是可能的.

下面进一步讨论LAR模型与NLAR模型的结构模型和预报模型的差异.

设$\{x_t\}$为平稳可逆LARMA(p,q)模型的解,即$\{x_t\}$满足如下随机差分方程

$$x_t = \varphi_1 x_{t-1} + \cdots + \varphi_p x_{t-p} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q}. \tag{5.1.10}$$

由性质1.2.3知,上式中的ε_t可唯一地表示为

$$\varepsilon_t = \sum_{j=0}^{\infty}\pi_j x_{t-j}, \tag{5.1.11}$$

式中系数$\pi_j(j=0,1,2,\cdots)$由(1.2.12)式确定.从而由(5.1.11)式,有

$$\begin{aligned}\hat{x}_n(1) &= E\{x_{n+1}\mid x_n, x_{n-1},\cdots\}\\ &= E\left\{\sum_{j=1}^{p}\varphi_j x_{n+1-j} + \varepsilon_{n+1} - \sum_{k=1}^{p}\theta_k\varepsilon_{n+1-k}\mid x_n, x_{n-1},\cdots\right\}\\ &= \varphi_1 x_n + \cdots + \varphi_p x_{n-p+1} - \theta_1\varepsilon_n - \cdots - \theta_q\varepsilon_{n-q+1}.\end{aligned}$$

由新息e_{n+1}的定义、(1.2.2)式和上式,有

$$e_{n+1} = x_{n+1} - \hat{x}_n(1) = \varepsilon_{n+1}. \tag{5.1.12}$$

(5.1.12)式表明了对平稳可逆LARMA模型(5.1.10),模型的新息序列与模型的白噪声序列是相同的.另一方面,如果将(5.1.10)式中的$\{\varepsilon_t\}$用平稳鞅差序列$\{e_t\}$代替时,由鞅差的定义及性质,仍可从(5.1.11)式得到

$$e_t = \sum_{j=0}^{\infty}\pi_j x_{t-j},$$

并且有

$$\hat{x}_n(1) = \varphi_1 x_n + \cdots + \varphi_p x_{n-p+1} - \theta_1 e_n - \cdots - \theta_q e_{n-q+1}.$$

从而由(5.1.2)式知,新息e_{n+1}就是平稳鞅差序列$\{e_t\}$在$n+1$时刻的元素e_{n+1}.上述讨论表明了对LARMA模型(5.1.10)如果假定$\{\varepsilon_t\}$为白噪声序列,则新息序列也是白噪声序列,且$e_t=\varepsilon_t$;如果$\{\varepsilon_t\}$为平稳鞅差序列,则新息序列也是平稳鞅差序列.如在§1.4所讨论的那样,从预报观点看,新息序列被假定为平稳鞅差序列是最合理的.所以在讨论LARMA模型时,从预报角度考虑,在模型(5.1.10)中用$\{e_t\}$代替$\{\varepsilon_t\}$更自然些.

但是,在NLARMA模型的情形,(5.1.12)式不一定成立,从而上述讨论对NLARMA模型也不一定适用.

例5.1.1 考虑如下NLAR模型

$$x_t = \frac{\varepsilon_t}{1 + x_{t-1}^2 + \cdots + x_{t-p}^2}, \tag{5.1.13}$$

其中$\{\varepsilon_t\}$为正态 $N(0,\sigma^2)$白噪声序列. 由定理 4.2.18 知,NLAR 模型(5.1.13)有唯一平稳解. 设$\{x_t\}$为(5.1.13)式的平稳解,那么

$$\begin{aligned}\hat{x}_n(1) &= E\{x_{n+1} | x_n, x_{n-1}, \cdots\} \\ &= E\left\{\frac{\varepsilon_{n+1}}{1 + x_n^2 + \cdots + x_{n-p+1}^2} \middle| x_n, x_{n-1}, \cdots\right\} \\ &= \frac{1}{1 + x_n^2 + \cdots + x_{n-p+1}^2} E\{\varepsilon_{n+1} | x_n, x_{n-1}, \cdots\} \\ &= 0.\end{aligned}$$

从而由(5.1.2)式,有

$$x_t = e_t. \tag{5.1.14}$$

特别地,

$$e_{n+1} = x_{n+1} - \hat{x}_n(1) = x_{n+1} = \frac{\varepsilon_{n+1}}{1 + x_n^2 + \cdots + x_{n-p+1}^2}.$$

由此可知,(5.1.12)式不能成立. 否则,必有

$$\varepsilon_{n+1} = \frac{\varepsilon_{n+1}}{1 + x_n^2 + \cdots + x_{n-p+1}^2}$$

成立. 由此导致 $1 + x_n^2 + \cdots + x_{n-p+1}^2 = 1$, a. s.. 于是有 $x_n = 0$, a. s. . 这与(5.1.14)式相矛盾.

通过例 5.1.1 不难看出线性序列与非线性序列在预报中的差异. 而且还可看到,对于(5.1.13)式的 NLAR 模型,在使用条件期望预报时却表现为 LAR(0)模型(5.1.14)式. 所以平稳序列$\{x_t\}$所满足的模型(5.1.13)式为$\{x_t\}$的结构模型,而(5.1.14)式为$\{x_t\}$的预报模型. 对于非线性时间序列,其结构模型和预报模型可以不相同. 由此可见,有必要用这两种不同的模型名称来加以区别. 而对线性序列的情况,如前所述,没有区别的必要.

现在讨论(5.1.6)呈现为无穷阶线性预报模型的情况,即

$$x_t = \sum_{j=1}^{\infty} \varphi_j x_{t-j} + e_t, \quad \sum_{j=1}^{\infty} \varphi_j^2 < \infty, \tag{5.1.15}$$

其中$\{e_t\}$为平稳鞅差序列. (5.1.15)式中的随机级数依概率收敛. 为叙述方便,仅限于讨论(5.1.6)式中的预报函数 φ 能表成(5.1.15)式右端的求和形式. 在§1.4 中,曾称(5.1.15)式为预报线性模型. 当(5.1.15)式中的 $e_t = \varepsilon_t$ 时,即 x_t 能表成如下的无穷阶 LAR 模型:

$$x_t = \sum_{j=1}^{\infty} \varphi_j x_{t-j} + \varepsilon_t. \tag{5.1.16}$$

与预报线性模型(5.1.15)相应，应当称(5.1.16)式为结构线性模型. 由§1.2中的讨论可知，线性平稳可逆 LARMA 模型(1.2.4)满足(1.2.11)式，即

$$x_t = \sum_{j=1}^{\infty} (-\pi_j) x_{t-j} + \varepsilon_t. \tag{5.1.17}$$

此式表明了平稳可逆 LARMA 模型属于结构线性模型. 进一步，平稳可逆 LARMA 模型的平稳解满足(1.2.7)式，即

$$x_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}. \tag{5.1.18}$$

此式表明了结构线性模型(5.1.17)的平稳解为线性序列. 由此联想到两个自然的问题是：结构线性模型在什么条件下有唯一的平稳解？其平稳解一定是形如(1.2.1)式的线性序列吗？

首先，为使(5.1.16)式有唯一的平稳解的确需要一些条件. 虽然给出结构线性模型(5.1.16)有唯一平稳解的充分必要条件是困难的，但是文献中有关于其充分条件的结果(见文献[58]、[125]、[133]). 在这些条件下，不但保证了(5.1.16)式有唯一平稳解，而且其解为(1.2.1)式的线性序列. 这些内容不是本书讨论的重点，故从略.

由于上述两个问题未彻底解决，于是随之而来的问题是：用(1.2.1)式定义的线性序列与(5.1.16)式的平稳解是否一致？如果不一致，当使用(1.2.1)式定义线性序列时，会使线性模型与线性序列相脱节. 如果相一致的话，则线性模型的解为线性序列；非线性模型的解为非线性序列. 由于上述是否一致的问题仍未解决，从而关于非线性序列的定义便存在一些不确定的因素. 到现在为止，我们只给出了§1.2关于线性序列的严格定义，这也是文献中常用的定义. 而对非线性序列和非线性模型没有给出严格的定义. 这些内容还有待研究.

下面讨论 LAR 模型和 NLAR 模型的另一差别. 这一差别表现在模型的多步预报方面. 为叙述方便，在下面论述中将结构模型简称模型.

首先考虑 LAR 模型

$$x_t = \varphi_1 x_{t-1} + \cdots + \varphi_p x_{t-p} + \varepsilon_t, \tag{5.1.19}$$

式中系数 $\varphi_1, \cdots, \varphi_p$ 满足条件(1.2.5)式，白噪声序列 $\{\varepsilon_t\}$ 满足(1.2.3)式. 由(5.1.19)式和(5.1.12)式的推导方法可得到

$$\hat{x}_n(1) = \varphi_1 x_n + \cdots + \varphi_p x_{n-p+1}. \tag{5.1.20}$$

再由(5.1.19)式还可得到

$$\begin{aligned}\hat{x}_n(l) &= E\{x_{n+l}\,|\,x_n,x_{n-1},\cdots\}\\ &= E\{\varphi_1 x_{n+l-1}+\cdots+\varphi_p x_{n+l-p}+\varepsilon_{n+l}\,|\,x_n,x_{n-1},\cdots\}\\ &= \varphi_1 E\{x_{n+l-1}\,|\,x_n,x_{n-1},\cdots\}+\cdots\\ &\quad+\varphi_p E\{x_{n+l-p}\,|\,x_n,x_{n-1},\cdots\}\\ &\quad+E\{\varepsilon_{n+l}\,|\,x_n,x_{n-1},\cdots\}\\ &= \varphi_1\hat{x}_n(l-1)+\cdots+\varphi_p\hat{x}_n(l-p).\end{aligned} \tag{5.1.21}$$

在(5.1.21)式中,当 $l-i\leqslant 0$ 时,记 $-k=l-i$.

$$\hat{x}_n(-k)=E\{x_{n-k}\,|\,x_n,x_{n-1},\cdots\}=x_{n-k}. \tag{5.1.22}$$

联合(5.1.21)和(5.1.22)两式,得到

$$\begin{cases}\hat{x}_n(l)=\varphi_1\hat{x}_n(l-1)+\cdots+\varphi_p\hat{x}_n(l-p);\\ \hat{x}_n(0)=x_n,\hat{x}_n(-1)=x_{n-1},\cdots,\hat{x}_n(-p+1)=x_{n-p+1}.\end{cases} \tag{5.1.23}$$

(5.1.23)式表明了 x_{n+l} 的多步预报满足 p 阶线性差分方程,可用递推方法逐一求出,计算十分方便. 此外,记 $\hat{\boldsymbol{x}}_n(l)=(\hat{x}_n(l),\cdots,\hat{x}_n(l-p+1))^\tau$,

$$\boldsymbol{\Phi}=\begin{pmatrix}\varphi_1 & \cdots & \varphi_{p-1} & \varphi_p\\ 1 & \cdots & 0 & 0\\ \vdots & \ddots & \vdots & \vdots\\ 0 & \cdots & 1 & 0\end{pmatrix},$$

则(5.1.23)式可写成

$$\hat{\boldsymbol{x}}_n(l)=\boldsymbol{\Phi}\hat{\boldsymbol{x}}_n(l-1). \tag{5.1.24}$$

由(5.1.24)式递推,可得到

$$\hat{\boldsymbol{x}}_n(l)=\boldsymbol{\Phi}^l\hat{\boldsymbol{x}}_n(0). \tag{5.1.25}$$

由(5.1.23)式可见,对 LAR 模型,不仅一步预报为 LAR 序列历史值的线性组合,而且多步预报也是 LAR 序列历史值的线性组合. 因此,(5.1.23)式中的 $\hat{x}_n(l)$既是条件期望预报,又是投影预报,两者是相等的. 对于 NLAR 模型,如前所述,两种预报不一定相等. 如果对非线性时间序列使用投影预报,将会如何呢?

考虑时间序列$\{x_t\}$,不妨设 $Ex_t=0$,用历史观测值 $x_n,x_{n-1},\cdots,x_{n-p+1}$ 对 x_{n+1}进行投影预报,并记为 $\hat{x}_n^*(1)$,则它应满足

$$E\{x_{n+1}-\hat{x}_n^*(1)\}^2=\inf_{c_1,\cdots,c_p}E\{x_{n+1}-\sum_{j=1}^{p}c_j x_{n-j+1}\}^2. \tag{5.1.26}$$

令 $\gamma_k=Ex_tx_{t-k},\ \boldsymbol{b}=(\gamma_1,\cdots,\gamma_p)^\tau,\ \boldsymbol{c}=(c_1,\cdots,c_p)^\tau,$

$$\Gamma = (\gamma_{i-j})_{1\leqslant i,j\leqslant p}.$$

由$\{x_t\}$的平稳性及 $Ex_t=0$，$Ex_t^2<\infty$，(5.1.26)式可表为

$$E\{x_{n+1}-\hat{x}_n^*(1)\}^2=\inf_c\{\gamma_0-c^\tau b+c^\tau\Gamma c\}. \qquad (5.1.27)$$

由时间序列分析中熟知的结果(见文献[125])，(5.1.27)式有唯一解：

$$\hat{x}_n^*(1)=b^\tau\Gamma^{-1}x_n, \qquad (5.1.28)$$

式中 $x_n=(x_n,\cdots,x_{n-p+1})^\tau$. 令 $\varphi=(\varphi_1,\cdots,\varphi_p)^\tau$，且

$$\varphi=\Gamma^{-1}b. \qquad (5.1.29)$$

当上述时间序列$\{x_t\}$满足 LAR(p)模型(5.1.19)时，(5.1.29)式恰好表达了其自回归系数与自协方差函数的关系，并被称为 Yule-Walker 方程(见文献[125]). 当$\{x_t\}$不满足(5.1.19)时，则可称(5.1.29)式是对$\{x_t\}$的一种近似模型描述，其系数仍由(5.1.29)式给出. 从而对$\{x_t\}$的多步预报可采用它的近似 LAR 模型的多步预报方法，也即用(5.1.23)式预报.

线性与非线性时间序列在预报方面的差异还表现在预报误差的方差上. 对于结构型 LAR 模型(5.1.19)，由(5.1.12)式，其预报误差为

$$e_{n+1}=\varepsilon_{n+1}.$$

从而

$$\begin{aligned}E\{e_{n+1}^2\mid x_n,x_{n-1},\cdots\}&=E\{\varepsilon_{n+1}^2\mid x_n,x_{n-1},\cdots\}\\&=E\varepsilon_{n+1}^2=\sigma^2.\end{aligned} \qquad (5.1.30)$$

但是，对于预报型 AR 模型(5.1.5)，无论它是线性的还是非线性的，由于(5.1.12)式不再成立，从而其一步预报误差的条件方差

$$E\{e_{n+1}^2\mid x_n,x_{n-1},\cdots\}=S^2(x_n,x_{n-1},\cdots) \qquad (5.1.31)$$

一般不为常数(而且上式右端的函数也不一定只依赖于 $x_n,\cdots,x_{n-p+1}$). 以模型(5.1.13)为例. 由(5.1.13)式和(5.1.14)式，有

$$\begin{aligned}E\{e_{n+1}^2\mid x_n,x_{n-1},\cdots\}&=E\{x_{n+1}^2\mid x_n,x_{n-1},\cdots\}\\&=\frac{1}{(1+x_n^2+\cdots+x_{n-p+1}^2)^2}E\{\varepsilon_{n+1}^2\mid x_n,x_{n-1},\cdots\}\\&=\frac{\sigma^2}{(1+x_n^2+\cdots+x_{n-p+1}^2)^2}.\end{aligned}$$

上式表明了 e_{n+1}的条件方差不为常数，从而不等于 Ee_{n+1}^2.

对于多步预报，在线性序列情况下，以模型(5.1.10)为例. 由(1.2.7)、(1.2.11)以及类似于求 $\hat{x}_n(1)$的方法，可得到

$$\hat{x}_n(l)=E\left\{\sum_{j=0}^{\infty}\psi_j\varepsilon_{n+l-j}\mid x_n,x_{n-1},\cdots\right\}$$

$$= \sum_{j=l}^{\infty} \psi_j \varepsilon_{n+l-j}.$$

从而 l 步预报的误差为

$$x_{n+l} - \hat{x}_n(l) = \sum_{j=0}^{l-1} \psi_j \varepsilon_{n+l-j}. \qquad (5.1.32)$$

于是,由(5.1.32)式有

$$\begin{aligned} &E\{[x_{n+l} - \hat{x}_n(l)]^2 \mid x_n, x_{n-1}, \cdots\} \\ &= E\{x_{n+l} - \hat{x}_n(l)\}^2 = E\left\{\sum_{j=0}^{l-1} \psi_j \varepsilon_{n+l-j}\right\}^2 \\ &= \sum_{j=0}^{l-1} \psi_j^2 E\varepsilon_{n+l-j}^2 = \sum_{j=0}^{l-1} \psi_j^2 \sigma^2 \quad (l \geqslant 1). \end{aligned} \qquad (5.1.33)$$

由(5.1.33)式可知,对线性序列的多步预报误差,其条件方差与无条件方差相等,且随预报步数增加而增加.

对非线性序列的多步预报,不仅其预报误差的条件方差与无条件方差不相等,而且存在着条件期望预报与投影预报两种不同的预报方法. 因此,其多步预报误差方差的特征也随其方法不同而不同. 在此仅叙述它们与线性序列预报的一同一异的特征.

考虑非线性平稳序列$\{x_t\}$,并用条件期望方法对其进行预报. 这时多步预报的误差方差总是随着预报步数而单调非降. 事实上,令

$$\sigma_l^2 = E[x_{n+l} - \hat{x}_n(l)]^2, \qquad (5.1.34)$$

由(5.1.1)和(5.1.6)式,有

$$x_{n+1} = \varphi(x_n, x_{n-1}, \cdots) + e_{n+1}. \qquad (5.1.35)$$

于是有

$$Ee_{n+1}^2 = \sigma_1^2, \qquad (5.1.36)$$

$$\begin{aligned} \hat{x}_n(2) &= E\{x_{n+2} \mid x_n, x_{n-1}, \cdots\} \\ &= E\{\varphi(x_{n+1}, x_n, x_{n-1}, \cdots) + e_{n+2} \mid x_n, x_{n-1}, \cdots\} \\ &= E\{\varphi(x_{n+1}, x_n, x_{n-1}, \cdots) \mid x_n, x_{n-1}, \cdots\} \\ &\triangleq \hat{\varphi}(x_{n+1}, x_n, x_{n-1}, \cdots), \end{aligned}$$

$$\begin{aligned} &E[x_{n+2} - \hat{x}_n(2)]^2 \\ &= E[\varphi(x_{n+1}, x_n, \cdots) + e_{n+2} - \hat{\varphi}(x_{n+1}, x_n, \cdots)]^2 \\ &= Ee_{n+2}^2 + E[\varphi(x_{n+1}, x_n, \cdots) - \hat{\varphi}(x_{n+1}, x_n, \cdots)]^2 \\ &\quad + 2Ee_{n+2}[\varphi(x_{n+1}, x_n, \cdots) - \hat{\varphi}(x_{n+1}, x_n, \cdots)]. \end{aligned} \qquad (5.1.37)$$

由于

$$
\begin{aligned}
& Ee_{n+2}[\varphi(x_{n+1},x_n,\cdots)-\hat{\varphi}(x_{n+1},x_n,\cdots)] \\
& \quad = E\{E\{e_{n+2}[\varphi(x_{n+1},x_n,\cdots) \\
& \qquad - \hat{\varphi}(x_{n+1},x_n,\cdots)] \mid x_n,x_{n-1},\cdots\}\} \\
& \quad = E\{[\varphi(x_{n+1},x_n,\cdots)-\hat{\varphi}(x_{n+1},x_n,\cdots)] \\
& \qquad \times E[e_{n+2} \mid x_n,x_{n-1},\cdots]\} \\
& \quad = 0,
\end{aligned} \tag{5.1.38}
$$

从而由(5.1.36)和(5.1.37)式得

$$\sigma_2^2 \geqslant \sigma_1^2. \tag{5.1.39}$$

对 $l>1$，记

$$
\begin{aligned}
\hat{x}_n(l) &= E\{x_{n+l} \mid x_n,x_{n-1},\cdots\} \triangleq \psi(x_n,x_{n-1},\cdots), \\
e_n(l) &= x_{n+l}-\hat{x}_n(l),
\end{aligned}
$$

于是再注意用平稳性有

$$Ee_n^2(l) = \sigma_l^2, \quad \forall\, n \geqslant 1, \tag{5.1.40}$$

$$E\{e_n(l) \mid x_n,x_{n-1},\cdots\} = 0, \text{a.s.}, \quad \forall\, n \geqslant 1, \tag{5.1.41}$$

$$x_{n+l+1} = \psi(x_{n+1},x_n,\cdots)+e_{n+1}(l), \quad \forall\, n \geqslant 1, \tag{5.1.42}$$

由(5.1.41)式有

$$
\begin{aligned}
\sigma_{l-1}^2 &= E[x_{n+l+1}-\hat{x}_n(l+1)]^2 \\
&= E[\psi(x_{n+1},x_n,\cdots)+e_{n+1}(l)-\hat{x}_n(l+1)]^2 \\
&= Ee_{n+1}^2(l)+E[\psi(x_{n-1},x_n,\cdots)-\hat{x}_n(l+1)]^2 \\
&\quad + 2Ee_{n+1}(l)[\psi(x_{n+1},x_n,\cdots)-\hat{x}_n(l+1)].
\end{aligned} \tag{5.1.43}
$$

类似于(5.1.38)式的推导，并注意使用(5.1.41)和(5.1.42)式可得

$$Ee_{n+1}(l)[\psi(x_{n+1},x_n,\cdots)-\hat{x}_n(l+1)] = 0, \tag{5.1.44}$$

从而联合(5.1.40)、(5.1.43)和(5.1.44)三式有

$$\sigma_{l+1}^2 \geqslant \sigma_l^2. \tag{5.1.45}$$

结合(5.1.39)和(5.1.45)两式，便证明了 σ_l^2 随 l 增加而非降. 这一特征与线性序列情况相同.

但是在非线性情况下，由于预报误差的条件方差不一定是常数，换句话说，即条件方差

$$E\{[x_{n+l}-\hat{x}_n(l)]^2 \mid x_n,x_{n-1},\cdots\} = S_l^2(x_n,x_{n-1},\cdots)$$

一般是 $(x_n,x_{n-1},\cdots)$ 的函数. 因此，它们不一定满足(5.1.45)式的单调性. 这就表现出非线性与线性序列在预报方面的又一差异. 下面的例子说

明了这种现象是存在的.

例 5.1.2　考虑如下双线性模型

$$x_t = \alpha x_{t-1}\varepsilon_t + \varepsilon_t, \tag{5.1.46}$$

其中$\{\varepsilon_t\}$是正态白噪声序列，$\varepsilon_t \sim N(0,1)$，且$|\alpha|<1$. 由 § 4.4 的讨论知，模型(5.1.46)有唯一平稳解. 又易见

$$\begin{aligned}\hat{x}_n(1) &= E\{x_{n+1}|x_n, x_{n-1}, \cdots\} \\ &= (\alpha x_n + 1)E\{\varepsilon_{n+1}|x_n, x_{n-1}, \cdots\} = 0.\end{aligned}$$

于是当 $l>1$ 时，

$$\begin{aligned}\hat{x}_n(l) &= E\{x_{n+l}|x_n, x_{n-1}, \cdots\} \\ &= E\{E\{x_{n+l}|x_{n+l-1}, x_{n+l-2}, \cdots\}|x_n, x_{n-1}, \cdots\} \\ &= 0.\end{aligned}$$

由上式和 $E\varepsilon_t^2=1$ 可得

$$\begin{aligned}S_n^2(l) &= E\{[x_{n+l} - \hat{x}_n(l)]^2|x_n, x_{n-1}, \cdots\} \\ &= E\{x_{n-l}^2|x_n, x_{n-1}, \cdots\} \\ &= E\{(\alpha x_{n+l-1} + 1)^2\varepsilon_{n+l}^2|x_n, x_{n-1}, \cdots\} \\ &= E\{E\{(\alpha x_{n+l-1} + 1)^2\varepsilon_{n+l}^2|x_{n+l-1}, \\ &\quad x_{n+l-2}, \cdots\}|x_n, x_{n-1}, \cdots\} \\ &= \alpha^2 S_n^2(l-1) + 1.\end{aligned} \tag{5.1.47}$$

由(5.1.47)式可得

$$\begin{aligned}S_n^2(1) &= (\alpha x_n + 1)^2, \\ S_n^2(2) &= \alpha^2(\alpha x_n + 1)^2 + 1.\end{aligned}$$

以上两式表明，当$(\alpha x_n+1)^2>1/(1-\alpha^2)$时，

$$S_n^2(2) < S_n^2(1). \tag{5.1.48}$$

另外，利用模型(5.1.46)不难验证，其平稳解的分布不是有界分布，从而$(\alpha x_n+1)^2>1/(1-\alpha^2)$以正概率出现. 这说明(5.1.48)式的不等式以正概率成立.

上述非线性序列的多步预报误差的条件方差的特征虽然比线性序列的情况复杂些，但却更合理地描述了许多客观现象. 不少客观存在的时间序列，其预报误差的条件方差是序列历史值的函数. 在某些情况下，可能出现类似于(5.1.48)式的现象，即预报步数增加，但预报误差的条件方差反而减少. 这一现象无疑是人们提出并重视 ARCH 模型的原因之一.

除本节所论及的线性与非线性序列预报的相同与相异的特征之外，

还有一些更为细致的差异，将在以后各节中加以介绍.

§ 5.2 可加噪声 NLAR 模型的预报

本节主要讨论形如(5.1.7)式的可加噪声 NLAR 模型的预报问题. 为了表达方便，仍使用(4.2.29)式的一阶高维模型形式，即

$$\boldsymbol{x}_t = \boldsymbol{\varphi}(\boldsymbol{x}_{t-1}) + \varepsilon_t \boldsymbol{u}. \tag{5.2.1}$$

在讨论预报问题时，总假定(5.2.1)式有唯一平稳解，且 $\boldsymbol{\varphi}$ 的形式和 ε_t 的分布已知. 在应用中，它们是可以通过估计得到的，或者是假定的，这里不加区别.

当模型(5.2.1)为 LAR 模型时，模型(5.2.1)的 l 步预报由(5.1.23)式确定. (5.1.23)式还表明了对 LAR 模型的 l 步预报，不需要使用 ε_t 的具体分布也能获得. 换言之，$\hat{x}_n(l)$ 可通过(5.1.24)和(5.1.25)式中的系数矩阵和历史值($x_n, x_{n-1}, \cdots, x_{n-p+1}$)而唯一确定.

当模型(5.2.1)为 NLAR 模型时，它的一步预报为

$$\begin{aligned}\hat{\boldsymbol{x}}_n(1) &= E\{\boldsymbol{\varphi}(\boldsymbol{x}_n) + \varepsilon_{n+1}\boldsymbol{u} \mid \boldsymbol{x}_n, \boldsymbol{x}_{n-1}, \cdots\} \\ &= \boldsymbol{\varphi}(\boldsymbol{x}_n) + \boldsymbol{u}E\{\varepsilon_{n+1} \mid \boldsymbol{x}_n, \boldsymbol{x}_{n-1}, \cdots\} \\ &= \boldsymbol{\varphi}(\boldsymbol{x}_n).\end{aligned} \tag{5.2.2}$$

其两步预报为

$$\begin{aligned}\hat{\boldsymbol{x}}_n(2) &= E\{\boldsymbol{x}_{n+2} \mid \boldsymbol{x}_n, \boldsymbol{x}_{n-1}, \cdots\} \\ &= E\{\boldsymbol{\varphi}(\boldsymbol{x}_{n+1}) + \varepsilon_{n+2}\boldsymbol{u} \mid \boldsymbol{x}_n, \boldsymbol{x}_{n-1}, \cdots\} \\ &= E\{\boldsymbol{\varphi}(\boldsymbol{\varphi}(\boldsymbol{x}_n) + \varepsilon_{n+1}\boldsymbol{u}) \\ &\quad + \varepsilon_{n+2}\boldsymbol{u} \mid \boldsymbol{x}_n, \boldsymbol{x}_{n-1}, \cdots\} \\ &= E\{\boldsymbol{\varphi}(\boldsymbol{\varphi}(\boldsymbol{x}_n) + \varepsilon_{n+1}\boldsymbol{u}) \mid \boldsymbol{x}_n\}.\end{aligned} \tag{5.2.3}$$

若 $\boldsymbol{x}_n = \boldsymbol{a}$，记 $\boldsymbol{\varphi}(\boldsymbol{a}) = \boldsymbol{b}$，则(5.2.3)式为

$$\hat{\boldsymbol{x}}_n(2) = E\boldsymbol{\varphi}(\boldsymbol{b} + \varepsilon_{n+1}\boldsymbol{u}). \tag{5.2.4}$$

(5.2.4)式中求均值是相对于 ε_{n+1} 的分布而言的，所以确定两步预报 $\hat{\boldsymbol{x}}_n(2)$ 的值不仅需要用到函数 $\boldsymbol{\varphi}$ 和 $\boldsymbol{x}_n = \boldsymbol{a}$ 的值，还要用到 ε_t 的分布. 这与线性情况不同. 由(5.2.3)和(5.2.4)的推导方法不难得到

$$\begin{aligned}\hat{\boldsymbol{x}}_n(l) &= E\{\boldsymbol{x}_{n+l} \mid \boldsymbol{x}_n = \boldsymbol{a}\} \\ &= E\{\boldsymbol{\varphi}(\boldsymbol{\varphi}(\cdots(\boldsymbol{\varphi}(\boldsymbol{\varphi}(\boldsymbol{a}) + \varepsilon_{n+1}\boldsymbol{u}) \\ &\quad + \varepsilon_{n+2}\boldsymbol{u})\cdots) + \varepsilon_{n+l-1}\boldsymbol{u})\}.\end{aligned} \tag{5.2.5}$$

上式中均值是用 $\varepsilon_{n+1},\cdots,\varepsilon_{n+l-1}$ 的联合分布计算的. 虽然 $\varepsilon_{n+1},\cdots,\varepsilon_{n+l-1}$ 是相互独立的随机变量,但计算(5.2.5)式仍是比较复杂的. 所以计算多步预报误差的方差和条件方差更加困难. 一般说来是没有明显公式的. 不过,一步预报却很特殊,由(5.2.1)和(5.2.2)两式可知,一步预报误差为 $\varepsilon_{n+1}\boldsymbol{u}$,从而其方差和条件方差相同,即

$$
\begin{aligned}
&E\{[\boldsymbol{x}_{n+1}-\hat{\boldsymbol{x}}_n(1)][\boldsymbol{x}_{n+1}-\hat{\boldsymbol{x}}_n(1)]^\tau|\boldsymbol{x}_n\} \\
&\quad = E\{\varepsilon_{n+1}^2\boldsymbol{u}\boldsymbol{u}^\tau|\boldsymbol{x}_n\} = \sigma^2\boldsymbol{u}\boldsymbol{u}^\tau.
\end{aligned}
$$

由 $\boldsymbol{x}_{n+1}$ 和 $\hat{\boldsymbol{x}}_n(1)$ 的定义及上式,有

$$
\begin{aligned}
&E\{[x_{n+1}-\hat{x}_n(1)]^2|x_n,x_{n-1},\cdots,x_{n-p+1}\} \\
&\quad = E\{\varepsilon_{n+1}^2|x_n,\cdots,x_{n-p+1}\} = \sigma^2. \qquad (5.2.6)
\end{aligned}
$$

在实际应用中,计算 $\hat{x}_n(l)$ $(l>1)$时,常采用蒙特卡洛方法. 具体地说,在已知函数 $\boldsymbol{\varphi}$ 、向量 $\boldsymbol{x}_n=\boldsymbol{a}$ 以及 ε_t 的分布时,用蒙特卡洛方法产生具有 ε_t 的分布的独立的伪随机数 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_N$,并计算如下平均值:

$$
\tilde{\boldsymbol{x}}_N(2) = \frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\varphi}\ (\boldsymbol{a}+\varepsilon_k\boldsymbol{u}), \qquad (5.2.7)
$$

基于熟知的大数定律,有

$$
\tilde{\boldsymbol{x}}_n(2) \to \hat{\boldsymbol{x}}_n(2), \quad 当\ N\to\infty.
$$

在实际应用中,可取 N 为适当大的正整数,以 $\tilde{\boldsymbol{x}}_N(2)$ 作为 $\hat{\boldsymbol{x}}_n(2)$ 的近似. 类似地,可以求 $\hat{\boldsymbol{x}}_n(l)$ 的近似值,例如

$$
\tilde{\boldsymbol{x}}_n(3) = \frac{1}{N^2}\sum_{k=1}^{N}\sum_{j=1}^{N}\boldsymbol{\varphi}\ (\boldsymbol{\varphi}\ (\boldsymbol{a}+\varepsilon_k\boldsymbol{u})+\varepsilon_j'\boldsymbol{u}), \qquad (5.2.8)
$$

式中 $\varepsilon_1,\cdots,\varepsilon_N$ 和 $\varepsilon_1',\cdots,\varepsilon_N'$ 为两组相互独立的、同分布的伪随机数列,且都与 ε_t 有相同的分布.

如果模型(5.2.1)中的 $\boldsymbol{\varphi}$ 已知,但是 ε_t 的分布未知,当有足够多的历史观测数据,即 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 已知时,则利用已知的 $\boldsymbol{\varphi}$ 函数、模型(5.2.1)和数据 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$,可计算出

$$
\varepsilon_k\boldsymbol{u} = \boldsymbol{x}_k - \boldsymbol{\varphi}\ (\boldsymbol{x}_{k-1}) \quad (k=1,2,\cdots,n). \qquad (5.2.9)
$$

由模型(5.2.1)的定义知,上述序列 $\varepsilon_k\boldsymbol{u}$ 是独立同分布的,从而可用 $\varepsilon_k\boldsymbol{u}$ 代替(5.2.7)式中的伪随机数列,所得到的 $\tilde{\boldsymbol{x}}_N(2)$ 同样是 $\hat{\boldsymbol{x}}_n(3)$ 的较好的近似. 为了计算 $\hat{\boldsymbol{x}}_n(3)$ 的近似值,可以使用统计中的重抽样方法. 例如从(5.2.9)式中的 n 个值按 $1/N$ 等概率进行随机抽样,先抽取 N 个 $\varepsilon_1',\cdots,\varepsilon_N'$,再独立地抽取 $\varepsilon_1'',\cdots,\varepsilon_N''$,以它们代入(5.2.8)式,便可得到 $\hat{\boldsymbol{x}}_n(3)$ 的近似值.

为了克服 NLAR 模型多步预报计算的困难，以及回避使用 ε_t 的具体分布，还有另外一些近似预报的方法.

第一种多步预报近似方法，是在§5.1 中所介绍的用 LAR 模型近似 NLAR 模型(5.2.1)的预报方法. 具体地说，在给定(5.2.1)式中的 φ 和 ε_t 的分布时，从理论上讲，平稳解 $\{x_t\}$ 的概率分布就给定了. 于是用(5.1.29)式可求出近似 LAR 模型的自回归系数向量，再用 LAR 模型的 l 步预报公式(5.1.25)，便可获得 $\{x_t\}$ 的 l 步预报. 它们可作为 $\hat{x}_n(l)$ 的近似预报，暂时将这一预报记为 $\hat{x}'_n(l)$.

第二种多步预报近似方法*)，是直接使用线性最小方差准则求如下的极小解，不妨记为 $\hat{x}''_n(l)$，即

$$E\{x_{n+l}-\hat{x}''_n(l)\}^2=\inf_{c_1,\cdots,c_p}E\Big\{x_{n+l}-\sum_{j=1}^{p}c_jx_{n-j+1}\Big\}^2. \quad (5.2.10)$$

不难看出，以上两种近似的 l 步预报都属于线性预报方法. 所以由(5.2.10)式可知

$$E\{x_{n+l}-\hat{x}'_n(l)\}^2\geqslant E\{x_{n+l}-\hat{x}''_n(l)\}^2. \quad (5.2.11)$$

由(5.1.25)式知，当模型(5.2.1)为 LAR(p)模型时，

$$\hat{x}_n(l)=\hat{x}'_n(l)=\hat{x}''_n(l), \quad (5.2.12)$$

从而(5.2.11)式中等式成立. 当模型(5.2.1)为 NLAR 模型时，(5.2.11)式中等式是否还成立? 回答是：(5.2.11)式中等式不一定成立.

例 5.2.1 考虑如下 NLAR(1)模型

$$x_t=\sin x_{t-1}+\varepsilon_t, \quad (5.2.13)$$

式中 $\{\varepsilon_t\}$ 为白噪声序列，ε_t 的分布有处处为正的密度，且满足

$$g(-x)=g(x)=o(|x|^{-4}),\ 当\ |x|\to\infty. \quad (5.2.14)$$

由例4.3.1知，模型(5.2.13)为几何遍历的. 由(5.2.13)和(5.2.14)式知，模型(5.2.13)的平稳解具有有穷的二阶矩. 下面先求模型(5.2.13)的预报. 由(5.2.2)式有

$$\hat{x}_n(1)=\sin x_n,$$

$$\sigma_1^2=E\{x_{n+1}-\hat{x}_n(1)\}^2=E\varepsilon_{n+1}^2=\sigma^2.$$

注意到 ε_t 的密度 g 是对称的，故有 $E\sin\varepsilon_t=0$. 再由(5.2.4)式有

$$\begin{aligned}\hat{x}_n(2)&=E\{\sin x_{n+1}|x_n\}\\&=E\{\sin(\sin x_n+\varepsilon_{n+1})\}\end{aligned}$$

*) 此方法是芝加哥大学刁锦寰教授在一次讨论班上提出的.

$$
\begin{aligned}
&= E\{\sin\sin x_n \cos\varepsilon_{n+1} + \cos\sin x_n \sin\varepsilon_{n+1}\} \\
&= (\sin\sin x_n) E\cos\varepsilon_{n+1}, \\
\sigma_2^2 &= E\{x_{n+2} - \hat{x}_n(2)\}^2 \\
&= E\{\sin(\sin x_n + \varepsilon_{n+1}) + \varepsilon_{n+2} \\
&\quad - (\sin\sin x_n) E\cos\varepsilon_{n+1}\}^2 \\
&= \sigma^2 + E(\sin\ \sin x_n)^2 (\cos\varepsilon_{n+1} - E\cos\varepsilon_1)^2 \\
&\quad + E(\cos\sin x_n \sin\varepsilon_{n+1})^2 \\
&\quad + 2E(\sin\sin x_n \cdot \cos\sin x_n) \\
&\quad \times E(\cos\varepsilon_{n+1} - E\cos\varepsilon_1)\sin\varepsilon_{n+1}. \qquad (5.2.15)
\end{aligned}
$$

注意到

$$
\begin{aligned}
2E(\cos\varepsilon_{n+1} - E\cos\varepsilon_1)\sin\varepsilon_{n+1} &= 2E\cos\varepsilon_{n+1}\sin\varepsilon_{n+1} \\
&= E\sin 2\varepsilon_{n+1} \\
&= 0,
\end{aligned}
$$

由(5.2.15)式便有

$$
\begin{aligned}
\sigma_2^2 = \sigma^2 &+ E(\sin\sin x_n)^2 [E\cos^2\varepsilon_{n+1} - (E\cos\varepsilon_1)^2] \\
&+ E(\cos\sin x_n)^2 E\sin^2\varepsilon_{n+1}.
\end{aligned}
$$

现在考虑对模型(5.2.13)使用第一种近似预报方法. 由(5.1.29)式可得

$$
\begin{aligned}
\varphi_1 &= (Ex_{n+1}^2)^{-1}(Ex_{n+1}x_n) \\
&= \frac{E(x_n \sin x_n)}{E(\sin^2 x_n) + \sigma^2},
\end{aligned}
$$

于是由(5.1.25)式，有

$$\hat{x}_n'(2) = \varphi_1^2 x_n. \qquad (5.2.16)$$

由第二种近似预报方法可得

$$\hat{x}_n''(2) = c_2 x_n, \qquad (5.2.17)$$

其中 $c_2 = (Ex_{n+2}^2)^{-1} E(x_{n+2}x_n)$. 不难验证

$$
\begin{aligned}
Ex_{n+2}^2 &= Ex_{n+1}^2 = E(\sin^2 x_n) + \sigma^2, \\
E(x_{n+2}x_n) &= E(x_n \sin\sin x_n)(E\cos\varepsilon_1). \qquad (5.2.18)
\end{aligned}
$$

记 ε_1 的特征函数为 $\psi_\varepsilon(\lambda)$，由(5.2.14)式有

$$
\begin{aligned}
\psi_\varepsilon(\lambda) &= \int_{-\infty}^{\infty} (\cos\lambda t + i\sin\lambda t) g(t) \mathrm{d}t. \\
&= \int_0^{\infty} 2\cos\lambda t g(t) \mathrm{d}t
\end{aligned}
$$

$$= E(\cos\lambda\varepsilon_1).$$

取 $\lambda=1$，有

$$E\cos\varepsilon_1 = \psi_\varepsilon(1).$$

另一方面，若$\{x_t\}$是模型(5.2.13)的平稳解，则$\{-x_t\}$必是

$$-x_t = \sin(-x_{t-1}) - \varepsilon_t$$

的平稳解. 由于 ε_t 与$-\varepsilon_t$ 同分布，故 x_t 与$-x_t$ 同分布，即模型(5.2.13)的不变分布是对称的.

利用上述事实，不难从满足(5.2.14)式的密度中找到满足如下条件的

$$Ex_n\sin x_n \neq 0,\quad E\cos\varepsilon_1 = \psi_\varepsilon(1) = 0. \tag{5.2.19}$$

由此式和(5.2.16)～(5.2.18)诸式有

$$\hat{x}_n'(2) = \varphi_1^2 x_n \neq 0,$$
$$\hat{x}_n''(2) = c_2 x_n = 0,$$
$$Ex_{n+2}x_n = 0.$$

进而又有

$$\begin{aligned}E\{x_{n+2} - \hat{x}_n'(2)\}^2 &= E\{x_{n+2} - \varphi_1^2 x_n\}^2\\ &= Ex_{n+2}^2 + \varphi_1^4 Ex_n^2 - 2\varphi_1^2 Ex_n x_{n+2}\\ &= Ex_{n+2}^2 + \varphi_1^4 Ex_n^2 > Ex_{n+2}^2\\ &= E\{x_{n+2} - \hat{x}_n''(2)\}^2.\end{aligned}$$

此例说明，在某些时候第二种近似方法的确比第一种近似方法好. 当然，它不比条件期望预报方法好. 对比这些方法的优劣，在实际应用中有重要意义. 将真实数据序列看作时间序列时，很少满足什么严格的 AR 模型. 所谓假定模型或者估计模型，都是对真实序列的一种近似描述. 换言之，"真实"的模型即使存在的话，也是非常复杂的，而使用模型，总是要适当地简化. 因此它们之间总有一定的出入. 在考虑预报方法时，应当使这种假定模型的出入的影响尽可能小. 这也就是统计学中的稳健统计的想法. 在这里不深入讨论这一概念，仅举一例说明之.

例 5.2.2 考虑平稳序列

$$x_t = \varepsilon_t + \varepsilon_{t-2}, \tag{5.2.20}$$

式中$\{\varepsilon_t\}$为正态白噪声序列，$\varepsilon_t \sim N(0,\sigma^2)$. 显然，(5.2.20)式中的$\{x_t\}$为平稳序列. 假如我们不知道它满足模型(5.2.20)，而认定它满足 AR(1) 模型. 换言之，认定的模型与真实的模型有出入. 此时，容易看出

$$Ex_t x_{t-1} = 0,$$

$$Ex_t x_{t-2} = \sigma^2 \neq 0.$$

于是使用前述的两种不同的近似预报方法，其结果是

$$\hat{x}'_n(2) = 0,$$

$$\hat{x}''_n(2) = \frac{Ex_t x_{t-2}}{Ex_t^2} x_n = \frac{1}{2} x_n.$$

由此可得 $E\{x_{n+2} - \hat{x}'_n(2)\}^2 = Ex_{n+2}^2 = 2\sigma^2$,

$$\begin{aligned} E\{x_{n+2} - \hat{x}''_n(2)\}^2 &= E\left\{\varepsilon_{n+2} + \varepsilon_n - \frac{1}{2}\varepsilon_n - \frac{1}{2}\varepsilon_{n-2}\right\}^2 \\ &= \sigma^2 + \frac{1}{4}\sigma^2 + \frac{1}{4}\sigma^2 \\ &= \frac{3}{2}\sigma^2 < E\{x_{n+2} - \hat{x}'_n(2)\}^2. \end{aligned}$$

上式表明，即使在线性情况下，当实际使用的模型失真时，(5.2.11)式也是成立的，而且严格不等式也可能成立.

总之，在考虑用线性方法作近似预报时，应首先考虑使用第二种近似方法. 无论对线性序列，还是对非线性序列，它都不比第一种近似方法差.

§ 5.3 随机条件方差模型的预报

考虑随机条件方差 NLAR 模型(4.2.24)，其相应的一阶高维模型为

$$\boldsymbol{x}_t = \boldsymbol{\varphi}(\boldsymbol{x}_{t-1}) + \varepsilon_t S(\boldsymbol{x}_{t-1})\boldsymbol{u}, \tag{5.3.1}$$

其中白噪声序列$\{\varepsilon_t\}$、变换 $\boldsymbol{\varphi}$ 和 S 如定理4.2.18或定理4.2.21的假定，这样，$\{\boldsymbol{x}_t\}$是几何遍历的，从而有唯一的平稳解.

类似于(5.2.2)和(5.2.6)的推导方法，可得到

$$\hat{\boldsymbol{x}}_n(1) = \boldsymbol{\varphi}(\boldsymbol{x}_{n-1}), \tag{5.3.2}$$

$$\begin{aligned} &E\{[\boldsymbol{x}_{n+1} - \hat{\boldsymbol{x}}_n(1)][\boldsymbol{x}_{n+1} - \hat{\boldsymbol{x}}_n(1)]^\tau | \boldsymbol{x}_n\} \\ &\quad = E\{\varepsilon_{n+1}^2 S^2(\boldsymbol{x}_n)\boldsymbol{u}\boldsymbol{u}^\tau | \boldsymbol{x}_n\} \\ &\quad = S^2(\boldsymbol{x}_n)\boldsymbol{u}\boldsymbol{u}^\tau, \end{aligned} \tag{5.3.3}$$

再由(5.3.3)式，有

$$\begin{aligned} &E\{[\boldsymbol{x}_{n+1} - \hat{\boldsymbol{x}}_n(1)][\boldsymbol{x}_{n+1} - \hat{\boldsymbol{x}}_n(1)]^\tau\} \\ &\quad = E\{E\{[\boldsymbol{x}_{n+1} - \hat{\boldsymbol{x}}_n(1)][\boldsymbol{x}_{n+1} - \hat{\boldsymbol{x}}_n(1)]^\tau | \boldsymbol{x}_n\}\} \\ &\quad = [ES^2(\boldsymbol{x}_n)]\boldsymbol{u}\boldsymbol{u}^\tau. \end{aligned} \tag{5.3.4}$$

由(5.3.3)和(5.3.4)两式可知,模型(5.3.1)的一步预报误差的条件方差不为常数,从而也不等于一步预报误差的无条件方差.

模型(5.3.1)的多步预报比§5.2的模型(5.2.1)更复杂.例如模型(5.3.1)的两步预报为

$$\begin{aligned}\hat{\boldsymbol{x}}_n(2) &= E\{\boldsymbol{x}_{n+2}\,|\,\boldsymbol{x}_n\}\\ &= E\{\boldsymbol{\varphi}(\boldsymbol{x}_{n+1}) + \varepsilon_{n+2}S(\boldsymbol{x}_{n+1})\boldsymbol{u}\,|\,\boldsymbol{x}_n\}\\ &= E\{\boldsymbol{\varphi}(\boldsymbol{x}_{n+1})\,|\,\boldsymbol{x}_n\}\\ &\quad + E\{E\{\varepsilon_{n+2}S(\boldsymbol{x}_{n+1})\boldsymbol{u}\,|\,\boldsymbol{x}_{n+1},\boldsymbol{x}_n\}\,|\,\boldsymbol{x}_n\}\\ &= E\{\boldsymbol{\varphi}(\boldsymbol{\varphi}(\boldsymbol{x}_n) + \varepsilon_{n+1}S(\boldsymbol{x}_n)\boldsymbol{u})\,|\,\boldsymbol{x}_n\}. \end{aligned} \tag{5.3.5}$$

若 $\boldsymbol{x}_n=\boldsymbol{a}$,则由(5.3.5)式得

$$\hat{\boldsymbol{x}}_n(2) = E\{\boldsymbol{\varphi}(\boldsymbol{\varphi}(\boldsymbol{a}) + \varepsilon_{n+1}S(\boldsymbol{a})\boldsymbol{u})\}. \tag{5.3.6}$$

当预报步数 $l>2$时,预报的表达式更加复杂,这里不再逐一写出了.

在考虑模型(5.3.1)的预报时,同§5.2一样,总是假定模型(5.3.1)中的函数 $\boldsymbol{\varphi}$ 和 S 以及 ε_t 的分布为已知的.即使如此,(5.3.6)式的 $\hat{\boldsymbol{x}}_n(2)$ 仍是难于计算的.§5.2所讨论的蒙特卡洛方法不难推广到(5.3.6)式的情形.

至于§5.2中提到的两种线性近似预报方法,对于模型(5.3.1)仍适用.由于§5.2中比较两种方法的讨论对一般平稳序列都适用,所以(5.2.11)式还成立.这里不再赘述.仅讨论以下例子:

例5.3.1 考虑如下非线性模型

$$x_t = \alpha x_{t-1} + \varepsilon_t(\beta_0 + \beta_1 x_{t-1}^{2\gamma})^{1/2}, \tag{5.3.7}$$

式中$|\alpha|<1$, $0\leqslant\gamma<1$, $\{\varepsilon_t\}$为正态白噪声序列,$\varepsilon_t\sim N(0,1)$, $x_t^{2\gamma}=|x_t|^{2\gamma}\geqslant 0$,$\beta_0\geqslant 0$,$\beta_1\geqslant 0$.由定理4.2.18知,模型(5.3.7)是几何遍历的.又由定理4.5.4知,模型(5.3.7)具有有穷的任意阶矩.对模型(5.3.7)的预报有如下结果:

$$\begin{aligned}\hat{x}_n(1) &= \alpha x_n,\\ E\{[x_{n+1} - \hat{x}_n(1)]^2\,|\,x_n\} &= E\{\varepsilon_{n+1}^2(\beta_0 + \beta_1 x_n^{2\gamma})\,|\,x_n\}\\ &= \beta_0 + \beta_1 x_n^{2\gamma},\\ \hat{x}_n(2) &= E\{x_{n+2}\,|\,x_n\}\\ &= E\{\alpha x_{n+1} + \varepsilon_{n+2}(\beta_0 + \beta_1 x_{n+1}^{2\gamma})^{1/2}\,|\,x_n\}\\ &= \alpha\hat{x}_n(1) = \alpha^2 x_n.\end{aligned}$$

实际上,若令

$$e_t = \varepsilon_t S(x_{t-1}),\ S(x_{t-1}) = (\beta_0 + \beta_1 x_{t-1}^{2\gamma})^{1/2},$$

易见，$\{e_t\}$是平稳鞅差序列，即

$$E\{e_t|x_{t-1},x_{t-2},\cdots\} = E\{e_t|x_{t-1}\} = 0.$$

从而(5.3.7)式具有形式

$$x_t = \alpha x_{t-1} + e_t. \tag{5.3.8}$$

按照§5.1中关于 LAR 模型预报的叙述知，模型(5.3.8)的多步预报满足

$$\hat{x}_n(l) = \alpha\hat{x}_n(l-1) = \cdots = \alpha^n x_n \quad (l = 1,2,\cdots). \tag{5.3.9}$$

以上预报的误差为

$$x_{n+l} - \hat{x}_n(l) = \alpha[x_{n+l} - \hat{x}_n(l-1)] + e_{n+l}.$$

利用$\{e_t\}$的鞅差性质，可知预报误差的条件方差满足

$$E\{[x_{n+l} - \hat{x}_n(l)]^2|x_n\} = \alpha^2 E\{[x_{n+l-1} - \hat{x}_n(l-1)]^2|x_n\} + E\{e_{n+l}^2|x_n\}. \tag{5.3.10}$$

记

$$\sigma_n^2(l) = E\{[x_{n+l} - \hat{x}_n(l)]^2|x_n\},$$
$$\delta_n(l) = E\{e_{n+l}^2|x_n\},$$

由(5.3.10)式有

$$\sigma_n^2(l) = \alpha^2\sigma_n^2(l-1) + \delta_n(l) \quad (l = 1,2,\cdots).$$

注意到 $\sigma_n^2(0)=0$，由上式可得

$$\sigma_n^2(l) = \delta_n(l) + \alpha^2\delta_n(l-1) + \cdots + \alpha^{2l-2}\delta_n(1). \tag{5.3.11}$$

再令

$$\sigma_l^2 = E[x_{n+l} - \hat{x}_n(l)]^2,$$
$$\begin{aligned}\sigma_1^2 &= Ee_{n+1}^2 \\ &= E\{E\{e_{n+1}^2|x_n\}\} \\ &= E(\beta_0 + \beta_1 x_n^{2\gamma}) \\ &= \beta_0 + \beta_1 Ex_n^{2\gamma}.\end{aligned}$$

由于模型(5.3.7)有唯一平稳解，故上式中诸量都与 n 无关. 再由$\{e_t\}$的鞅差性质及(5.3.10)得

$$\sigma_l^2 = \alpha^2\sigma_{l-1}^2 + \sigma_1^2, \tag{5.3.12}$$

从而有

$$\begin{aligned}\sigma_l^2 &= \sigma_1^2 + \alpha^2\sigma_1^2 + \cdots + \alpha^{2l-2}\sigma_1^2 \\ &= \frac{1-\alpha^{2l}}{1-\alpha^2}\sigma_1^2.\end{aligned} \tag{5.3.13}$$

由(5.3.13)式可知,l 步预报误差的无条件方差是随 l 增加而增加的.但是,其条件方差不一定随 l 增加而增加.这一特征对非线性序列的预报均适用,并且在§5.1中已论证过.对于具体的模型(5.3.7)而言,可用线性差分方程(5.3.9)和(5.3.13)式分别以递推方式计算前 l 步预报和预报误差的无条件方差.

以上的讨论还可以推广到如下的模型,即

$$x_t = \alpha x_{t-1} + \cdots + \alpha_p x_{t-p} + \varepsilon_t(\beta_0 + \beta_1 x_{t-1}^{2\gamma} + \cdots + \beta_q x_{t-q}^{2\gamma})^{1/2}, \tag{5.3.14}$$

其中自回归系数 $\alpha=(\alpha_1,\cdots,\alpha_p)^\tau$ 满足条件(1.2.5)式, $0\leqslant\gamma\leqslant1$, $\beta_0>0$, $\beta_i\geqslant0$ $(i=1,2,\cdots,q)$, $\{\varepsilon_t\}$ 为正态白噪声序列, $\varepsilon_t\sim N(0,1)$.由定理4.2.18知,模型(5.3.14)是几何遍历的,从而有唯一的平稳解.记

$$e_t = \varepsilon_t(\beta_0 + \beta_1 x_{t-1}^{2\gamma} + \cdots + \beta_q x_{t-q}^{2\gamma})^{1/2}, \tag{5.3.15}$$

则$\{e_t\}$是平稳鞅差序列.从而(5.3.14)式又可写为

$$x_t = \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + e_t. \tag{5.3.16}$$

按照LAR模型的 l 步预报的求法,得

$$\hat{x}_n(l) = \alpha_1 \hat{x}_n(l-1) + \cdots + \alpha_p \hat{x}_n(l-p) \quad (l=1,2,\cdots), \tag{5.3.17}$$

而且

$$\begin{aligned}
\hat{x}_n(1) &= E\{x_{n+1}|x_n,\cdots,x_{n-p+1}\} \\
&= E\{\alpha_1 x_n + \cdots + \alpha_p x_{n-p+1} + e_{n+1}|x_n,\cdots,x_{n-p+1}\} \\
&= \alpha_1 x_n + \cdots + \alpha_p x_{n-p+1} + E\{e_{n+1}|x_n,\cdots,x_{n-p+1}\} \\
&= \alpha_1 x_n + \cdots + \alpha_p x_{n-p+1} \\
&\quad + E\{E\{e_{n+1}|x_n,x_{n-1},\cdots\}|x_n,\cdots,x_{n-p+1}\} \\
&= \alpha_1 x_n + \cdots + \alpha_p x_{n-p+1} \\
&= E\{x_{n+1}|x_n,x_{n-1},\cdots\}.
\end{aligned}$$

类似地有

$$\begin{aligned}
\hat{x}_n(l) &= E\{x_{n+l}|x_n,\cdots,x_{n-p+1}\} \\
&= E\{x_{n+l}|x_n,x_{n-1},\cdots\}.
\end{aligned}$$

但是

$$\begin{aligned}
&S^2(x_n,x_{n-1},\cdots) \\
&= E\{e_{n+1}^2|x_n,x_{n-1},\cdots\} \\
&= E\{[x_{n+1} - \hat{x}_n(1)]^2|x_n,x_{n-1},\cdots\}
\end{aligned}$$

$$
\begin{aligned}
&= E\{\varepsilon_{n+1}^2(\beta_0 + \beta_1 x_n^{2\gamma} + \cdots + \beta_q x_{n-q+1}^{2\gamma}) \mid x_n, x_{n-1}, \cdots\} \\
&= \beta_0 + \beta_1 x_n^{2\gamma} + \cdots + \beta_q x_{n-q+1}^{2\gamma} \\
&= E\{e_{n+1}^2 \mid x_n, \cdots, x_{n-q+1}\}.
\end{aligned} \tag{5.3.18}
$$

(5.3.18)式表明了除非 $p>q$，否则，上式右端的 $S^2(x_n, x_{n-1}, \cdots)$ 不是 $(x_n, \cdots, x_{n-p+1})$ 的函数. 这也是预报 LAR(p)模型与结构 LAR(p)模型的不同之处.

关于模型(5.3.14)的预报误差方差，可借助于预报 LAR(p)模型(5.3.16)进行讨论. 在平稳条件下，其 l 步预报误差具有类似于(5.1.32)式的形式，即

$$
x_{n+l} - \hat{x}_n(l) = \sum_{j=0}^{l-1} \psi_j e_{n+l-j}.
$$

利用 $\{e_t\}$ 的平稳鞅差性质和(5.3.15)式，可得

$$
\begin{aligned}
\sigma_l^2 &= E[x_{n+l} - \hat{x}_n(l)]^2 = \sum_{j=0}^{l-1} \psi_j^2 E e_t^2 \\
&= \left(\sum_{j=0}^{l-1} \psi_j^2\right) E[\varepsilon_t^2(\beta_0 + \beta_1 x_{t-1}^{2\gamma} + \cdots + \beta_q x_{t-q}^{2\gamma})] \\
&= \left(\sum_{j=0}^{l-1} \psi_j^2\right)(\beta_0 + \beta_1 E x_{t-1}^{2\gamma} + \cdots + \beta_q E x_{t-q}^{2\gamma}) \\
&= \left\{\beta_0 + \left(\sum_{j=1}^{q} \beta_j\right)(E x_t^{2\gamma})\right\} \sum_{j=0}^{l-1} \psi_j^2.
\end{aligned}
$$

但是，关于 $e_{n+l} = x_{n+l} - \hat{x}_n(l)$ 的条件方差，一般不易得到比较简明的结果. 当然，对于 p 取较小的值时，例如 $p=2$ 或 $p=3$ 时，仍可直接导出 $\sigma_n^2(l)$ 的递推公式. 这里从略.

在时间序列预报中，还应当考虑区间预报. 由于非线性序列的预报比较复杂，其预报误差的分布更复杂，所以其区间预报也就更困难. 只有在比较特殊的情况下，才能容易地给出区间预报. 例如对§5.2中可加噪声 NLAR 模型(5.1.5)的一步预报的区间估计问题，由于此时一步预报的误差为 ε_{n+1}，在 ε_t 的分布给定时，给出 x_{n+1} 的区间估计——区间预报，与线性序列的区间预报没有区别. 另一种特殊情况是本节例5.3.1及其推广情形. 为了方便于讨论区间估计，特别假定 $\{\varepsilon_t\}$ 为标准正态白噪声序列.

首先讨论例5.3.1的一步区间预报问题. 此时，一步预报的误差为

$$
e_{n+1} = \varepsilon_{n+1}(\beta_0 + \beta_1 x_n^{2\gamma})^{1/2}. \tag{5.3.19}
$$

注意到 $\varepsilon_{n+1} \sim N(0,I)$，$e_{n+1}$的条件分布为 $N(0,\beta_0+\beta_1 x_n^{2\gamma})$，记为

$$e_{n+1} | x_n \sim N(0,\beta_0 + \beta_1 x_n^{2\gamma}). \qquad (5.3.20)$$

以置信水平 $\alpha=0.01$为例，由

$$P(|e_{n+1}| \leqslant 1.96(\beta_0 + \beta_1 x_n^{2\gamma})^{1/2} | x_n) = 0.01$$

和

$$x_{n+1} = \alpha x_n + e_{n+1}$$

可知，区间预报为

$$(\alpha x_n - 1.96(\beta_0 + \beta_1 x_n^{2\gamma})^{1/2}, \quad \alpha x_n + 1.96(\beta_0 + \beta_1 x_n^{2\gamma})^{1/2}),$$

即未来值 x_{n+1}落在以上区间的条件概率为99%.

由(5.3.7)和(5.3.9)两式有

$$\hat{x}_n(2) = \alpha^2 x_n,$$

$$\begin{aligned} x_{n+2} &= \alpha x_{n+1} + \varepsilon_{n+2}(\beta_0 + \beta_1 x_{n+1}^{2\gamma})^{1/2} \\ &= \alpha^2 x_n + \alpha\varepsilon_{n+1}(\beta_0 + \beta_1 x_n^{2\gamma})^{1/2} \\ &\quad + \varepsilon_{n+2}\{\beta_0 + \beta_1[\alpha x_n + \varepsilon_{n+1}(\beta_0 + \beta_1 x_n^{2\gamma})^{1/2}]^{2\gamma}\}^{1/2}. \end{aligned}$$

由上式可知，两步预报的误差为

$$\begin{aligned} x_{n+2} - \hat{x}_n(2) &= \alpha\varepsilon_{n+1}(\beta_0 + \beta_1 x_n^{2\gamma})^{1/2} \\ &\quad + \varepsilon_{n+2}\{\beta_0 + \beta_1[\alpha x_n + \varepsilon_{n+1}(\beta_0 + \beta_1 x_n^{2\gamma})^{1/2}]^{2\gamma}\}^{1/2}. \qquad (5.3.21) \end{aligned}$$

虽然假定$\{\varepsilon_t\}$为标准正态白噪声序列，但两步预报的误差(5.3.21)式的分布仍是很复杂的. 由此可以想象，对于比(5.3.7)更复杂的模型(5.3.14)，要解决其多步预报的区间估计问题，只能借助于各种近似方法. 这些还有待进一步研究.

最后我们指出，对于随机条件方差模型，其预报误差的条件方差是随机的. 由此可知，其区间预报的区间长度也是随机的. 正如(5.3.20)式所表示的，这一事实告诉我们，在某些条件下，对未来的预报可能较粗糙；而在另一些条件下，则可能较精确，这在实际应用中是很有意义的.

§ 5.4 非线性 ARMA 模型的预报

在线性时间序列预报中，平稳可逆 LARMA 模型的预报理论和方法被研究得较多，应用也非常广泛. 在平稳可逆 LARMA 序列$\{x_t\}$的预报中，可逆性条件保证了(1.2.11)式成立，从而有如下的重要事实：

$$\mathscr{F}_n^x = \sigma(x_t, t \leqslant n) = \mathscr{F}_n^\varepsilon = \sigma(\varepsilon_t, t \leqslant n), \qquad (5.4.1)$$

式中 $\sigma(x_t,t\leqslant n)$表示由$\{x_t,t\leqslant n\}$生成的 σ-代数. $\sigma(\varepsilon_t,t\leqslant n)$有类似的含

义.(5.4.1)式在讨论 LARMA(p,q)模型(1.2.4)的预报时起着重要作用.利用(5.4.1)和(1.2.11)两式可得到如下模型:

$$x_{n+1}=\alpha_1 x_n+\cdots+\alpha_p x_{n-p+1}+\varepsilon_{n+1}-\theta_1\varepsilon_n-\cdots-\theta_q\varepsilon_{n-q+1} \tag{5.4.2}$$

的一步预报

$$\hat{x}_n(1)=\alpha_1 x_n+\cdots+\alpha_p x_{n-p+1}-\theta_1\varepsilon_n-\cdots-\theta_q\varepsilon_{n-q+1}, \tag{5.4.3}$$

式中 $\varepsilon_n,\cdots,\varepsilon_{n-q+1}$ 虽然不能被直接量测,但是可以利用 $x_n,x_{n-1},\cdots$ 和(1.2.11)式计算,至少在理论上是如此.进一步还可给出 l 步预报,以及预报误差的方差(与条件方差相等).这些都是线性时间序列分析中熟知的内容(如见文献[125]).

但是对非线性 ARMA 模型,如在§4.6中所述,具有可逆性的模型并不多,而且也很难验证其可逆性.因此,一般说来,(5.4.1)式对非线性 ARMA 序列并不成立.这给非线性 ARMA 模型的预报带来了比以前两节更多的困难.例如,考虑如下的 NLARMA(1,1)模型

$$x_t=\psi(x_{t-1},\varepsilon_{t-1})+\varepsilon_t, \tag{5.4.4}$$

其一步预报为

$$\begin{aligned}\hat{x}_n(1)&=E\{\psi(x_n,\varepsilon_n)+\varepsilon_{n+1}\,|\,x_n,x_{n-1},\cdots\}\\&=E\{\psi(x_n,\varepsilon_n)\,|\,x_n,x_{n-1},\cdots\}.\end{aligned} \tag{5.4.5}$$

在(5.4.5)式中,虽然二元函数 ψ 已知,历史值 $x_n,x_{n-1},\cdots$ 已知,但是由于 ε_n 是不可直接量测的,又不知 ε_n 是否为 $\mathscr{F}_n^x$ 可测的,所以(5.4.5)式中 $\psi(x_n,\varepsilon_n)$ 的条件期望是很难有简单表达式的,除非模型(5.4.4)具有可逆性,即 ε_t 为 $\mathscr{F}_n^x$ 可测的.

现在讨论如下的双线性模型:

$$x_t=\sum_{j=1}^{p}\alpha_j x_{t-j}+\sum_{j=1}^{q}\theta_j\varepsilon_{t-j}+\sum_{i=1}^{P}\sum_{j=1}^{Q}\beta_{ij}x_{t-i}\varepsilon_{t-j}+\varepsilon_t. \tag{5.4.6}$$

在适当的条件下(见§4.6),模型(5.4.6)是平稳可逆的,这里不讨论这些内容了.模型(5.4.6)的特例的可逆性在§4.6中已有讨论.下面讨论在平稳可逆时,模型(5.4.6)的预报问题.由(5.4.1)和(5.4.6)可知,一步预报为

$$\hat{x}_n(1)=E\{x_{n+1}\,|\,x_n,x_{n-1},\cdots\}$$

$$= \sum_{j=1}^{p}\alpha_j x_{n+1-j} + \sum_{j=1}^{q}\theta_j\varepsilon_{n+1-j} + \sum_{i=1}^{P}\sum_{j=1}^{Q}\beta_{ij}x_{n+1-i}\varepsilon_{n+1-j}. \tag{5.4.7}$$

类似地，两步预报为

$$\begin{aligned}\hat{x}_n(2) = {} & \alpha_1\hat{x}_n(1) + \sum_{j=2}^{p}\alpha_j x_{n+2-j} + \sum_{j=2}^{q}\theta_j\varepsilon_{n+2-j} \\ & + \sum_{i=2}^{P}\sum_{j=2}^{Q}\beta_{ij}x_{n+2-i}\varepsilon_{n+2-j} \\ & + \sum_{j=2}^{Q}\beta_{ij}\hat{x}_n(1)\varepsilon_{n+2-j} \\ & + \beta_{11}E\{x_{n+1}\varepsilon_{n+1}|x_n,x_{n-1},\cdots\} \\ = {} & \Big(\alpha_1 + \sum_{j=2}^{Q}\beta_{ij}\varepsilon_{n+2-j}\Big)\hat{x}_n(1) + \sum_{j=2}^{P}\alpha_j x_{n+2-j} \\ & + \sum_{j=2}^{q}\theta_j\varepsilon_{n+2-j} + \sum_{i=2}^{P}\sum_{j=2}^{Q}\beta_{ij}x_{n+2-i}\varepsilon_{n+2-j} \\ & + \beta_{11}E\{x_{n+1}\varepsilon_{n+1}|x_n,x_{n-1},\cdots\}.\end{aligned} \tag{5.4.8}$$

由(5.4.6)式，取 $t=n+1$，并在其两端同乘以 ε_{n+1}，再求条件期望，可得

$$\begin{aligned} & E\{x_{n+1}\varepsilon_{n+1}|x_n,x_{n-1},\cdots\} \\ & = E\{[\sum_{j=1}^{p}\alpha_j x_{n+1-j} + \sum_{j=1}^{q}\theta_j\varepsilon_{n+1-j} \\ & \quad + \sum_{i=1}^{P}\sum_{j=1}^{Q}\beta_{ij}x_{n+1-i}\varepsilon_{n+1-j} + \varepsilon_{n+1}]\varepsilon_{n+1}|x_n,x_{n-1},\cdots\} \\ & = E\{\varepsilon_{n+1}^2|x_n,x_{n-1},\cdots\} \\ & = E\varepsilon_{n+1}^2 \\ & = \sigma^2. \end{aligned} \tag{5.4.9}$$

将(5.4.9)式代入(5.4.8)式可得 $\hat{x}_n(2)$. 对于更多步的预报 $\hat{x}_n(l)$ 也可按上述方法求得，这里从略.

在上述讨论中，假定了模型(5.4.6)具有可逆性，却没有给出 ε_n 用怎样的$(x_n,x_{n-1},\cdots)$的函数表达出来. 实际上，这也是很难给出的. 在应用中，在模型可逆的前提下，$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 可用如下递推方法近似求出，即由(5.4.6)式可得

$$\varepsilon_t = x_t - \sum_{j=1}^{p}\alpha_j x_{t-j} - \sum_{j=1}^{q}\theta_j \varepsilon_{t-j}$$
$$- \sum_{i=1}^{P}\sum_{j=1}^{Q}\beta_{ij}x_{t-i}\varepsilon_{t-j} \quad (t = 1,2,\cdots,n), \tag{5.4.10}$$

式中的初始值 $\varepsilon_0,\varepsilon_{-1},\cdots,\varepsilon_{-m}$ 和 $x_0,x_{-1},\cdots,x_{-l}$ 可用零值代替，其中 $m=\max(q,Q),l=\max(p,P)$. 当 n 较大时，$\varepsilon_n,\varepsilon_{n-1},\cdots$等的值与真值近似得较好. 这种近似方法是仿效了 LARMA 模型预报方法. 最后讨论一个简单双线性模型.

例 5.4.1　考虑如下双线性模型：

$$x_t = \alpha x_{t-1} + \beta x_{t-1}\varepsilon_{t-2} + \varepsilon_t, \tag{5.4.11}$$

式中$\{\varepsilon_t\}$为正态白噪声序列，$\varepsilon_t \sim N(0,\sigma^2)$，且 $\alpha^2+\beta^2\sigma^2<1$. 在此条件下，由 § 4.4和 § 4.5中的讨论知，模型(5.4.11)有平稳解，且平稳解具有有穷的二阶矩. 由(5.4.11)式，有

$$\begin{aligned}\varepsilon_t &= (x_t - \alpha x_{t-1}) + \beta x_{t-1}\varepsilon_{t-2} \\ &= (x_t - \alpha x_{t-1}) + \beta x_{t-1}(x_{t-2} - \alpha x_{t-3}) \\ &\quad + \beta^2 x_{t-1}x_{t-3}\varepsilon_{t-4} \\ &= \cdots\cdots \\ &= (x_t - \alpha x_{t-1}) + \sum_{k=1}^{\infty}\beta^k(x_{t-2k} - \alpha x_{t-2k-1}) \\ &\quad \times \prod_{j=1}^{k}x_{t-2j+1}. \end{aligned} \tag{5.4.12}$$

于是模型(5.4.11)的一步预报为

$$\begin{aligned}\hat{x}_n(1) &= E\{x_{n+1}|x_n,x_{n-1},\cdots\} \\ &= \alpha x_n + \beta x_n\varepsilon_{n-1} \\ &= (\alpha + \beta\varepsilon_{n-1})x_n,\end{aligned}$$

两步预报为

$$\begin{aligned}\hat{x}_n(2) &= E\{x_{n+2}|x_n,x_{n-1},\cdots\} \\ &= \alpha\hat{x}_n(1) + \beta\hat{x}_n(1)\varepsilon_n \\ &= (\alpha + \beta\varepsilon_n)\hat{x}_n(1). \\ &= (\alpha + \beta\varepsilon_n)(\alpha + \beta\varepsilon_{n-1})x_n.\end{aligned}$$

它们的误差的条件方差分别为

$$\begin{aligned}&E\{[x_{n+1} - \hat{x}_n(1)]^2|x_n,x_{n-1},\cdots\} \\ &= E\{\varepsilon_{n+1}^2|x_n,x_{n-1},\cdots\}\end{aligned}$$

$$
\begin{aligned}
&= E\varepsilon_{n+1}^2 \\
&= \sigma^2,
\end{aligned}
$$

$$
\begin{aligned}
&E\{[x_{n+2} - \hat{x}_n(2)]^2 \mid x_n, x_{n-1}, \cdots\} \\
&\quad = E\{(\alpha + \beta\varepsilon_n)^2\varepsilon_{n+1}^2 + \varepsilon_{n+2}^2 \mid x_n, x_{n-1}, \cdots\} \\
&\quad = \sigma^2(\alpha + \beta\varepsilon_n)^2 + \sigma^2.
\end{aligned}
$$

由以上两式可见，对双线性模型(5.4.11)，其一步预报误差的条件方差是常数，但两步预报误差的条件方差则是 ε_n 的函数，而 ε_n 又可用(5.4.12)式表成 $x_n, x_{n-1}, \cdots$ 的函数.

第 6 章

时间序列的线性性检验

根据观测数据序列，判断其是否为某一线性时间序列的样本序列，属于时间序列统计分析中的线性性检验问题. 这在时间序列分析中具有重要意义. 本章将介绍时间序列分析中线性性检验问题的分类和某些检验方法，以及在非线性时间序列分析中的作用.

§6.1 检验的分类

线性时间序列的结构理论及模型估计和预报方法，比非线性时间序列要简便得多. 因此，在分析真实数据序列之前，首先应当检验它们是否为线性序列的样本. 确切地说，检验它们的非线性是否显著. 如果不显著，则说明这些数据是线性序列的样本，或者可近似地看作是线性序列的样本. 于是应当按线性时间序列分析的理论与方法分析该数据序列. 如果不进行线性性检验，而盲目地按非线性时间序列分析的理论与方法去分析真实数据序列，当真实数据序列确为(或者近似为)线性序列时，则会事倍功半，得不偿失. 这不仅由于非线性时间序列分析的计算量很大，而且其统计分析的精度也不如线性情况好. 所以说，除非在统计分析真实数据序列之前已经知道该序列为线性序列，或者为非线性序列，一般应当先进行线性性检验.

为了讨论时间序列的线性性检验问题，首先需要明确“线性性”的含义. 如像前几章中所介绍的，平稳序列有预报线性性和结构线性性两种. 我们曾以模型(5.1.13)为例来说明了两者不是相同的概念. 于是，时间序列的线性性检验就要分为预报线性性与结构线性性两种检验问题. 我们仍以模型(5.1.13)为例来说明两种检验的区别.

考虑非线性模型(5.1.13)，即

$$x_t = \frac{\varepsilon_t}{1 + x_{t-1}^2 + \cdots + x_{t-p}^2}, \tag{6.1.1}$$

并且 $p\geqslant 1$. 按照§5.1的分析知,上述序列满足预报 LAR(0)模型,即

$$x_t = e_t,$$

从而$\{x_t\}$是预报线性的. 如果一种检验预报线性性的方法具有相容性,当样本长度 n 趋于无穷大时,则判断上述$\{x_t\}$为预报线性序列的概率趋于1. 但是,一种检验结构线性性的方法,如果也具有相容性,当样本长度 n 趋于无穷大时,则判断上述$\{x_t\}$为结构线性序列的概率趋于零.

为了进一步阐明上述两种检验的区别,再考虑如下的较特殊的假设检验问题,即根据样本 $x_1,\cdots,x_n$,检验如下的对立假设:

H_0:$\{x_t\}$为如下的预报 LAR(p)序列

$$x_t = \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + e_t, \tag{6.1.2}$$

其中系数 $\boldsymbol{\alpha}=(\alpha_1,\cdots,\alpha_p)^\tau$ 满足条件(1.2.5)式;

$$H_1:\{x_t\}\text{ 不是(6.1.2)式的预报 LAR}(p)\text{ 序列}. \tag{6.1.3}$$

注意,H_0 成立的充分必要条件是:(6.1.2)式中的$\{e_t\}$为平稳鞅差序列. 当模型(6.1.2)的阶数和系数已知时,由样本序列 $x_1,\cdots,x_n$ 可得到$\{e_t\}$的样本序列 $e_{p+1},\cdots,e_n$. 于是上述的假设(6.1.2)和(6.1.3)就等价于下面的假设:

$$H_0:e_{p+1},\cdots,e_n\text{ 为平稳鞅差序列的样本}; \tag{6.1.4}$$

$$H_1:H_0\text{ 不真}. \tag{6.1.5}$$

如果在上述假设检验问题中,将检验预报 LAR 模型改为检验结构 LAR 模型,根据类似的理由,我们得到等价的检验问题是:

$$H_0:e_{p+1},\cdots,e_n\text{ 为白噪声序列的样本}; \tag{6.1.6}$$

$$H_1:H_0\text{ 不真}. \tag{6.1.7}$$

现在分别考虑上述两种不同的问题. 首先讨论预报线性性检验问题,仍沿用第5章的记号,记

$$\hat{x}_n(1) = E\{x_{n+1}|x_n,x_{n-1},\cdots\} = \varphi(x_n,x_{n-1},\cdots), \tag{6.1.8}$$

称$\{x_t\}$为预报线性的,如果(6.1.8)中的 $\varphi(x_n,x_{n-1},\cdots)$满足

$$\hat{x}_n^*(1) = \hat{E}\{x_{n+1}|x_n,x_{n-1},\cdots\} = \varphi(x_n,x_{n-1},\cdots). \tag{6.1.9}$$

在讨论预报线性性检验时,我们面临的第一个问题是:(6.1.9)中的函数 φ 可能是无穷多元的函数,而真实数据总是有限的. 为解决这一矛盾,我们先考虑(6.1.9)式中的 φ 为有穷多元函数,例如 p 元函数 $\varphi(x_n,\cdots,x_{n-p+1})$. 在实际应用中,$p$ 可以取适当大,或者用文献中的 AIC 准则(见文献[8])估计 p 的值. 此估计值可以随 n 的增大而增大,从而可视为对

无穷多元函数 φ 的逼近. 当限定 p 值有穷时，并且已给定或者被估计出来时，便将预报线性性检验问题转化为(6.1.2)和(6.1.3)的检验问题. 此时，我们又面临第二个问题，即在(6.1.2)式中系数是未知的. 为了解决这一问题，可以根据样本 $x_1,\cdots,x_n$，估计(6.1.2)式中的未知参数 $\boldsymbol{\alpha}$，从而得到 $\hat{\boldsymbol{\alpha}}$，其估计方法可用 LARMA 模型统计分析中熟知的方法(如见文献[125]、[20]和[133]). 于是可得到

$$\hat{e}_k = x_k - \hat{a}_1 x_{k-1} - \cdots - \hat{a}_p x_{k-p} \quad (k = p+1,\cdots,n). \tag{6.1.10}$$

这些 $\hat{e}_k$ 可作为(6.1.2)式中 e_k 的估计值. 这时，我们面临的第三个问题是：要检验对立假设(6.1.2)和(6.1.3)等价于检验对立假设(6.1.4)和(6.1.5)，其中需检验 e_t 是否为平稳鞅差序列，而我们只有 $\hat{e}_{p+1},\cdots,\hat{e}_n$. 为解决这一矛盾，可借助理论分析，在适当的条件下，检验 e_t 是否为平稳鞅差与检验 $\hat{e}_t$ 是等价的.

解决上述问题是检验预报线性性所使用的重要近似手段. 当然，也有使用其他近似方法的，例如假定 e_t 为白噪声或为正态白噪声. 但是使用这些假定，便会将检验预报线性性改为检验结构线性性，从而混淆了两种不同的概念.

与预报线性性检验相类似，为解决结构线性性检验，也必须采用同样的近似手段. 如第 5 章所述，迄今为止，对于结构线性模型和线性序列的定义尚无定式，再加上对结构模型的检验将会导致对 e_t 为白噪声序列的检验，这些都造成解决如下对立假设检验问题的困难，即

$$H_0: \text{序列为结构线性的}; \tag{6.1.11}$$

$$H_1: H_0 \text{ 不真}. \tag{6.1.12}$$

在文献中，也未见到讨论上述检验问题的论述. 代替上述的完全对立假设检验问题，以 H_1 为某个指定的非线性序列，如指定为双线性型的非线性序列. 更具体些，可将 H_1 限定为某个参数型的非线性模型，如

$$x_t = \sum_{j=1}^{p} \alpha_j x_{t-j} + \sum_{i=1}^{r}\sum_{j=1}^{s} \beta_{ij} x_{t-i}\varepsilon_{t-j} + \varepsilon_t. \tag{6.1.13}$$

于是，对(6.1.13)式的线性性检验就等价于如下的检验：

$$H_0: \beta_{ij} = 0 \quad (i=1,\cdots,r;\ j=1,\cdots,s); \tag{6.1.14}$$

$$H_1: \sum_{i,j} |\beta_{ij}| \neq 0. \tag{6.1.15}$$

在文献中，除了以上的结构线性性检验的近似方法以外，还有用高阶

谱方法. 利用线性序列的高阶谱特有的性质建立检验方法. 毫无疑问，也需要采用适当的近似手段，这里不详述了.

§ 6.2 Kolmogorov-Smirnov 检验方法

在这一节里，我们讨论预报 LAR(p)模型的检验问题，并假定阶数 p 是已知的. 在实际应用中，阶数 p 往往是未知的，这可用 § 6.1 中提到的 AIC 准则进行估计，并以估计值代替未知的 p 值. 在理论分析中，当阶数 p 被估计值代替时，或者 p 为无穷时，而以适当大的 p 值作近似，将会遇到许多困难或较深的研究课题. 本书不打算涉及太多太烦琐的数学内容，所以仅限于介绍阶数已知的 AR(p)模型的线性性检验问题.

考虑平稳时间序列$\{x_t\}$，其一步预报仅与前 p 个值有关，即

$$\begin{aligned} E\{x_t \mid x_{t-1}, x_{t-2}, \cdots\} &= E\{x_t \mid x_{t-1}, \cdots, x_{t-p}\} \\ &= \varphi(x_{t-1}, \cdots, x_{t-p}). \\ &= \varphi(\boldsymbol{x}_{t-1}), \end{aligned} \tag{6.2.1}$$

其中 $\boldsymbol{x}_t = (x_t, \cdots, x_{t-p+1})^\tau$，$\varphi$ 为某个 p 元可测函数. 记

$$e_t = x_t - \varphi(x_{t-1}, \cdots, x_{t-p}),$$

则有

$$x_t = \varphi(x_{t-1}, \cdots, x_{t-p}) + e_t, \tag{6.2.2}$$

其中$\{e_t\}$为平稳鞅差序列，即

$$E\{e_t \mid x_{t-1}, \cdots, x_{t-p}\} = 0, \text{ a.s.}. \tag{6.2.3}$$

所谓模型(6.2.2)的线性性检验问题，就是检验函数 φ 是否为线性函数，或者说，检验 φ 的非线性性是否显著. 当 φ 为线性函数时，(6.2.2)式为

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + e_t. \tag{6.2.4}$$

上述的检验问题可以表述为如下的对立假设检验问题：

H_0：φ 是线性的；

H_1：φ 不是线性的.

在文献中已有许多文章讨论这一检验问题. 其方法可分为三类：一类是近似的参数化检验方法，也就是说，用某一类指定的有限参数型的非线性函数类代替上述的 H_1 假设(如见文献[75]、[100]). 第二类是非参数化方法，即直接检验上述的完全对立的假设检验问题. 它是基于对模型(6.2.2)中未知函数 φ 的非参数估计，以及对线性模型(6.2.4)的拟合(见文献[66]). 第三类只是基于线性预报新息序列分析的方法. 此方法也是

检验完全对立的假设检验问题，不过它是基于对线性模型(6.2.4)的拟合残差分析(见文献[5]和[118]).

在上述各类检验方法中，尤其是第二类的非参数化方法，当模型(6.2.2)中的阶数 p 较大时，可能会遇到统计学中的"维数灾难"问题. 在此，我们简要地叙述"维数灾难"一词的含义. 这一术语是在研究高维数据分析中提出来的. 例如在高维非线性回归分析中就有此"灾难性"的困难. 简单地说，对于非线性非参数化的回归模型

$$y_t = \varphi(\boldsymbol{x}_t) + e_t \quad (t = 1, \cdots, n), \tag{6.2.5}$$

式中 $\boldsymbol{x}_t = (x_{1t}, \cdots, x_{pt})^{\tau}$，无论采用何种非参数化方法估计未知函数 φ，本质上都是基于 φ 函数在各点的邻域加权平均. 为了说明"维数灾难"现象，不妨说满足(6.2.5)式的 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$ 是 p 维单位立方体中的均匀样本. 要估计 φ 在 $\boldsymbol{x}$ 点的值 $\varphi(\boldsymbol{x})$，计算落入以 $\boldsymbol{x}$ 为中心的、边长为 δ 的 p 维立方体内的 $\boldsymbol{x}_i$ 的平均值 $\hat{\varphi}(\boldsymbol{x})$，以 $\hat{\varphi}(\boldsymbol{x})$ 作为 $\varphi(\boldsymbol{x})$ 的估计. 例如当 $p=1, n=1000$ 时，只要取 δ 适当小，例如取 $\delta = 1/100$，假定 φ 足够光滑，则 φ 在 $(x-1/200, x+1/200)$ 内可近似为常数. 由于假定 $x_1, \cdots, x_n$ 均匀散布在 $(0,1)$ 内，从而在 $(x-1/200, x+1/200)$ 中大约有 $n \times \frac{1}{100} = 10$ 个数据. 按照上述同样的考虑，当 p 较大时，例如 $p=10$，那么，边长为 1/100 的 10 维立方体的体积只有 $(1/100)^{10}$. 若 n 个数据均匀散布在边长为 1 的 10 维立方体中，则落在以 $\boldsymbol{x}$ 为中心，边长为 1/100 的 10 维立方体中的数据大约有 $\left(\frac{1}{100}\right)^{10} \times n = n \cdot 10^{-20}$ 个. 若要保证其中的数据不少于 10 个，则数据个数要大于 10^{21} 个. 这一过分的要求，源于 p 维邻域的体积与邻域半径的 p 次方成正比. 对于较大的 p 而言，用邻域加权平均的想法解决非参数回归分析遇到了难以克服的难题. 这就被称为"维数灾难". 这一现象在非线性自回归分析中同样存在. 所以说，建立在对模型(6.2.2)中函数 φ 进行非参数估计基础上的检验，必然会遇到"维数灾难"问题. 这一问题在本书第 7 章中还将提到.

现在简要讨论检验(6.2.2)式中 φ 的线性性的三类方法的优缺点：在第一类方法中，当真函数 φ 属于被指定的非线性函数类时，其检验功效较好；当真函数 φ 不属于指定的函数类时，其功效较差，甚至于很差. 或者说，此方法不具有相容性. 另外，当 p 较大时，其计算量将增加，但是不涉及"维数灾难"问题. 第二类方法不限定 φ 的具体函数类. 所以，在适当条件下检验具有相容性. 但是，由于使用了非参数自回归估计方法，从而遇

到了"维数灾难"问题. 换言之,此类方法只适用于较低阶 p 的情况. 在第三类方法中,也不限定 φ 的类型,因此在适当条件下,可望得到检验的相容性质. 但是在此类方法中,有的方法——本节要介绍的 Kolmogorov-Smirnov 检验方法,虽然未涉及非参数自回归估计问题,但仍遇到"维数灾难"的麻烦. 而在下一节中介绍的混合型检验方法既不涉及非参数估计,也没有"维数灾难"的麻烦. 但是,此方法以及第三类的其他方法,一般说来其功效较差. 特别是当对 φ 的类型有确切了解时,应当采用第一类检验方法.

下面叙述 Kolmogorov-Smirnov 检验(以下简称 K-S 检验)方法的基本原理.

对平稳序列 $\{x_t\}$,考虑一步最佳线性预报,即求如下的最小解问题:

$$\inf_{\beta_0,\beta_1,\cdots,\beta_p} E\{x_t - \beta_0 - \beta_1 x_{t-1} - \cdots - \beta_p x_{t-p}\}^2$$
$$= E[x_t - \varphi_0 - \varphi_1 x_{t-1} - \cdots - \varphi_p x_{t-p}]^2. \tag{6.2.6}$$

记

$$\xi_t = x_t - \varphi_0 - \varphi_1 x_{t-1} - \cdots - \varphi_p x_{t-p}, \tag{6.2.7}$$

无论 $\{x_t\}$ 是否为 LAR(p) 序列,ξ_t 总是满足(6.2.6)式的 p 阶线性最小方差预报的误差. 当 $\{x_t\}$ 的确满足(6.2.4)式的 LAR 模型时,则

$$\begin{aligned}
&E\{x_t - \beta_0 - \beta_1 x_{t-1} - \cdots - \beta_p x_{t-p}\}^2 \\
&= E[e_t + \alpha_0 + \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} \\
&\quad - \beta_0 - \beta_1 x_{t-1} - \cdots - \beta_p x_{t-p}\}^2 \\
&= Ee_t^2 + 2E\{(\alpha_0 - \beta_0) + (\alpha_1 - \beta_1)x_{t-1} + \cdots \\
&\quad + (\alpha_p - \beta_p)x_{t-p}\}e_t \\
&\quad + E\{(\alpha_0 - \beta_0) + (\alpha_1 - \beta_1)x_{t-1} + \cdots \\
&\quad + (\alpha_p - \beta_p)x_{t-p}\}^2 \\
&= Ee_t^2 + E\{(\alpha_0 - \beta_0) + (\alpha_1 - \beta_1)x_{t-1} + \cdots \\
&\quad + (\alpha_p - \beta_p)x_{t-p}\}^2 \\
&\geqslant Ee_t^2.
\end{aligned}$$

在上式中利用了 $\{e_t\}$ 满足(6.2.3)式的性质. 易见,上式中等号成立的条件是 $\alpha_i=\beta_i \quad (i=0,1,\cdots,p)$. 换言之,当 $\{x_t\}$ 满足(6.2.4)式的线性模型时,其系数恰好是(6.2.6)式最佳线性预报的系数解,即 $\varphi_i=\alpha_i \quad (i=0,1,\cdots,p)$. 从而由(6.2.7)式定义的 ξ_t 满足

$$\xi_t = e_t,$$

即 ξ_t 为满足(6.2.3)式的鞅差序列.

如果$\{x_t\}$满足 NLAR 模型(6.2.2),仍设 $\varphi_0,\varphi_1,\cdots,\varphi_p$ 为满足(6.2.6)式的解. 此时由(6.2.7)式定义的 ξ_t 不再满足(6.2.3)式. 因为如果 ξ_t 满足(6.2.3)式,则有

$$\begin{aligned}0 &= E\{\xi_t \mid x_{t-1},\cdots,x_{t-p}\}\\ &= E\{x_t - \varphi_0 - \varphi_1 x_{t-1} - \cdots - \varphi_p x_{t-p} \mid x_{t-1},\cdots,x_{t-p}\}\\ &= E\{e_t + \varphi(x_{t-1},\cdots,x_{t-p})\\ &\qquad - \varphi_0 - \varphi_1 x_{t-1} - \cdots - \varphi_p x_{t-p} \mid x_{t-1},\cdots,x_{t-p}\}\\ &= \varphi(x_{t-1},\cdots,x_{t-p}) - \varphi_0 - \varphi_1 x_{t-1} - \cdots - \varphi_p x_{t-p},\end{aligned}$$

此即

$$\varphi(x_{t-1},\cdots,x_{t-p}) = \varphi_0 + \varphi_1 x_{t-1} + \cdots + \varphi_p x_{t-p}.$$

从而表明 φ 是线性的,这与上述前提相矛盾!

综合以上两方面的分析知,φ 为线性的充分必要条件是:由(6.2.6)和(6.2.7)式定义的$\{\xi_t\}$为鞅差序列,即满足(6.2.3)式. 进一步,由概率测度积分的知识知,ξ_t 满足(6.2.3)式的充分必要条件是

$$E\xi_t I(\boldsymbol{x}_{t-1} < \boldsymbol{s}) = 0,\quad \forall\, \boldsymbol{s} \in R^p, \tag{6.2.8}$$

式中 $\boldsymbol{x}_t = (x_t,\cdots,x_{t-p+1})^\tau$, $\boldsymbol{s} = (s_1,\cdots,s_p)^\tau$,

$$I(\boldsymbol{x}_{t-1} < \boldsymbol{s}) = I(x_{t-1} < s_1)\cdot\cdots\cdot I(x_{t-p} < s_p). \tag{6.2.9}$$

显然,(6.2.8)式等价于

$$\sup_s |E\xi_t I(\boldsymbol{x}_{t-1} < s)| = 0. \tag{6.2.10}$$

换言之,$\{x_t\}$满足 LAR 模型(6.2.4)式的充分必要条件是(6.2.10)式成立.

重新审查上述的推理分析还可发现:为确定(6.2.10)式中的 ξ_t,需要知道(6.2.7)式中的 $\varphi_0,\varphi_1,\cdots,\varphi_p$,而由(6.2.6)式确定这些系数时,需要知道$(x_t,\cdots,x_{t-p})$的联合分布. 我们的目的是检验$\{x_t\}$满足 LAR 模型(6.2.4)还是满足 NLAR 模型(6.2.2),当然不会知道$(x_t,\cdots,x_{t-p})$的联合分布. 所以说上述的推理只是解决检验问题的理论基础. 为了得到可以使用的检验方法,还需要引入以下的具体步骤,主要包括:完成满足(6.2.6)式的线性拟合;求出拟合残差;寻找基于(6.2.10)式的检验统计量;并寻找该统计量的渐近分布;从而解决(6.2.2)式的线性性检验问题. 下面以 $p=1$ 为例,叙述具体计算的内容和步骤.

第一步:利用样本 $x_1,\cdots,x_n$,求解一阶自回归系数的最小二乘估计,

即求解如下极小值问题：

$$\inf_{\beta_0,\beta_1}\sum_{t=2}^{n}(x_t-\beta_0-\beta_1x_{t-1})^2$$

$$=\sum_{t=2}^{n}(x_t-\hat{\varphi}_0-\hat{\varphi}_1x_{t-1})^2.$$

计算拟合残差

$$\hat{\xi}_t=x_t-\hat{\varphi}_0-\hat{\varphi}_1x_{t-1}\quad(t=2,3,\cdots,n)$$

和拟合残差方差

$$\hat{\sigma}^2=\frac{1}{n}\sum_{t=2}^{n}\hat{\xi}_t^2.$$

第二步：取 $m_n=[n/\log n]$，其中$[x]$表示取 x 的整数部分. 计算

$$S_n(a)=\frac{1}{m_n^{1/2}\hat{\sigma}}\sum_{t=2}^{m_n}\hat{\xi}_tI(x_{t-1}<a),\tag{6.2.11}$$

$$S_n=\sup_a|S_n(a)|.\tag{6.2.12}$$

第三步：给出 S_n 的渐近分布. 由于这要涉及较繁琐的数学论证，在此我们只介绍其结果，有兴趣的读者可查阅文献[5].

设 x_t 的分布为 $F(x)$，利用 F 的单调性可得到

$$S_n(a)=\frac{1}{m_n^{1/2}\hat{\sigma}}\sum_{t=2}^{m_n}\hat{\xi}_tI(F(x_{t-1})<F(a))$$

$$=\frac{1}{m_n^{1/2}\hat{\sigma}}\sum_{t=2}^{m_n}\hat{\xi}_tI(u_{t-1}<F(a))$$

$$\triangleq\eta_n(F(a)),$$

其中 $u_{t-1}=F(x_{t-1})$. 当 F 连续时，u_t 为(0,1)上均匀分布. 令 $b=F(a)\in[0,1]$，于是当 H_1 成立时，有

$$S_n=\sup_a|S_n(a)|=\sup_b|\eta_n(b)|\to\infty,\ \text{a.s.}.\tag{6.2.13}$$

当 H_0 成立时，有

$$\eta_n(b)\Rightarrow W(b),\tag{6.2.14}$$

其中“⇒”表示弱收敛，$W(b)$为$[0,1]$上的标准布朗运动，从而

$$S_n=\sup_{0\leqslant b\leqslant 1}|\eta_n(b)|\xrightarrow{d}\eta=\sup_{0\leqslant b\leqslant 1}|W(b)|,\tag{6.2.15}$$

其中“$\xrightarrow{d}$”表示依分布收敛. 随机变量 η 的分布在文献中已给出(见文献[5]).

第四步：对给定的置信水平 β，求出 W_β，使得

$$P(\eta > W_\beta) = \beta. \qquad (6.2.16)$$

当(6.2.15)式中的 $S_n > W_\beta$ 时，则拒绝 H_0，即认为 φ 非线性显著.

当 $p>1$ 时，推广上述检验方法遇到本质性困难，困难主要来源于上述第三步. 此时(6.2.13)式中的 $S_n(a)$ 为向量 $\boldsymbol{a}=(a_1,\cdots,a_p)^\tau$ 的函数，即

$$\begin{aligned} S_n(\boldsymbol{a}) &= \frac{1}{m_n^{1/2}\hat{\sigma}} \sum_{t=p+1}^{m_n} \hat{\xi}_t I(\boldsymbol{x}_{t-1} < \boldsymbol{a}) \\ &= \frac{1}{m_n^{1/2}\hat{\sigma}} \sum_{t=p+1}^{m_n} \hat{\xi}_t I(x_{t-1} < a_1) \cdot \cdots \cdot I(x_{t-p} < a_p) \\ &= \frac{1}{m_n^{1/2}\hat{\sigma}} \sum_{t=p+1}^{m_n} \hat{\xi}_t I(F(x_{t-1} < F(a_1)) \cdot \cdots \\ &\quad \cdot I(F(x_{t-p}) < F(a_p)) \\ &= \frac{1}{m_n^{1/2}\hat{\sigma}} \sum_{t=p+1}^{m_n} \hat{\xi}_t I(u_{t-1} < b_1) \cdot \cdots \cdot I(u_{t-p} < b_p) \\ &\triangleq \frac{1}{m_n^{1/2}\hat{\sigma}} \sum_{t=p+1}^{m_n} \hat{\xi}_t I(\boldsymbol{u}_{t-1} < \boldsymbol{b}) \\ &\triangleq \eta_n(\boldsymbol{b}), \end{aligned} \qquad (6.2.17)$$

其中 $\boldsymbol{u}_{t-1}=(u_{t-1},\cdots,u_{t-p})^\tau, \boldsymbol{b}=(b_1,\cdots,b_p)^\tau, b_i=F(a_i) \ (i=1,\cdots,p)$. 当 $n\to\infty$ 时，在 H_1 成立时，(6.2.13)式仍成立. (6.2.14)式也可推广到 $p>1$ 的情况. 但是(6.2.14)式右边的 $W(\boldsymbol{b})$ 是多指标布朗运动，而且它的分布结构与 e_t 的分布有关. 这与 $p=1$ 的情况有实质性差别. 在 $p=1$ 时，(6.2.14)式右端的 $W(b)$ 为标准布朗运动，它的分布结构与 e_t 的分布无关. 在统计学中，称 $\eta_n(b)$ 的极限分布独立于 e_t 的分布. 在 $p>1$ 时，$\eta_n(\boldsymbol{b})$ 的极限分布依赖于 e_t 的分布. 这种差别导致了无法给出(6.2.15)式的极限分布. 况且从概率论的基础知识来看，即使(6.2.14)式中的 $W(\boldsymbol{b})$ 为 p 指标的标准布朗运动，当 $n\to\infty$ 时，其极值

$$S_n = \sup_{0\leq b_1,\cdots,b_p\leq 1} |\eta_n(\boldsymbol{b})| \xrightarrow{d} \sup_{0\leq b_1,\cdots,b_p\leq 1} |W(\boldsymbol{b})| = \eta. \qquad (6.2.18)$$

但是 η 的分布也未被研究者所获得. 只是在 $p=1$ 时，η 的分布才有可以利用的结果.

为了使用 K-S 检验方法解决 $p>1$ 时的线性性检验问题，我们介绍如下的近似方法：

第一步：利用样本 $x_1,\cdots,x_n$，求解 p 阶自回归模型的最小二乘估计，即求如下极小值解：

$$\inf_{\beta_0,\beta_1,\cdots,\beta_p}\sum_{t=p+1}^{n}\{x_t-\beta_0-\beta_1x_{t-1}-\cdots-\beta_px_{t-p}\}^2$$

$$=\sum_{t=p+1}^{n}\{x_t-\hat{\varphi}_0-\hat{\varphi}_1x_{t-1}-\cdots-\hat{\varphi}_px_{t-p}\}^2. \tag{6.2.19}$$

计算拟合残差

$$\hat{\xi}_t=x_t-\hat{\varphi}_0-\hat{\varphi}_1x_{t-1}-\cdots-\hat{\varphi}_px_{t-p}\quad(t=p+1,\cdots,n), \tag{6.2.20}$$

和拟合残差方差

$$\hat{\sigma}^2=\frac{1}{n-p}\sum_{t=p+1}^{n}\hat{\xi}_t^2. \tag{6.2.21}$$

第二步：取 $m_n=[n/\log n]$，计算

$$S_{ni}(a)=\frac{1}{m_n^{1/2}\hat{\sigma}}\sum_{t=p+1}^{m_n}\hat{\xi}_tI(x_{t-i}<a)$$

$$=\frac{1}{m_n^{1/2}\hat{\sigma}}\sum_{t=p+1}^{m_n}\hat{\xi}_tI(F(x_{t-i})<F(a))$$

$$\triangleq\eta_{ni}(F(a)), \tag{6.2.22}$$

$$S_{ni}=\sup_{0\leqslant b\leqslant 1}|\eta_{ni}(b)|, \tag{6.2.23}$$

$$S_n=\max\{S_{n1},\cdots,S_{np}\}. \tag{6.2.24}$$

第三步：与 $p=1$ 时的第四步相同，从略.

以上的方法只是检验(6.2.2)线性性的一种近似检验方法. 此时(6.2.15)式对每个 $\hat{\eta}_{ni}$成立，即

$$S_{ni}\xrightarrow{d}\eta\quad(i=1,\cdots,p).$$

但是当(6.2.2)式中的 φ 为非线性时，S_n 并不满足(6.2.13)式. 况且即使在 H_0 成立时，S_n 的渐近分布也是很复杂的. 所以上述方法仅是一种近似.

在此顺便指出，根据上述构造 K-S 检验统计量的基本原理，可以设计检验(3.3.1)式中的条件方差 S^2 是否为常数的方法，从而决定是否需要使用 § 3.3 中的随机条件方差模型，详见文献[24].

§ 6.3　混合型检验方法

本节要介绍的方法，仍是为解决检验模型(6.2.2)式是否为线性模型

的问题.与前一节的方法一样,这也是一种非参数型的完全对立的假设检验,而且也是基于线性模型(6.2.4)的拟合残差的统计运算.与 K-S 检验方法的不同之处是:本节的方法不是用单一类型的检验统计量,而是基于两种不同类型的统计量之和进行检验.因此,这一检验方法称为混合型检验方法.它是从非线性回归检验引申而来的(见文献[118]).

现在介绍混合型方法的理论依据,首先由 § 6.2 的讨论知,H_0 成立的充分必要条件是(6.2.10)式成立.但是基于这一特征所建立的 K-S 检验统计量 S_n(见(6.2.17)、(6.2.18)式),当 $p\geqslant 2$ 时,其渐近分布与 e_t 的分布有关,且不知其具体关系,而且当 p 较大时,计算 S_n 时还会遇到"维数灾难"的问题.这些困难使得我们不能把前一节的检验方法有效地推广到$p>1$的情况.

现在注意(6.2.10)式成立,当且仅当对任何非负权函数

$$\int\{E\xi_t I(\boldsymbol{x}_{t-1}<\boldsymbol{a})\}^2 W_2(\boldsymbol{a})\mathrm{d}\boldsymbol{a}=0 \tag{6.3.1}$$

成立.利用这一特征,将(6.3.1)式中的 $E\xi_t I(\boldsymbol{x}_{t-1}<\boldsymbol{a})$用(6.2.11)式中的 $S_n(\boldsymbol{a})$代替,则可得到一个统计量

$$v_n=c_n\int\{S_n(\boldsymbol{a})\}^2 W_2(\boldsymbol{a})\mathrm{d}\boldsymbol{a}, \tag{6.3.2}$$

式中 c_n 为适当选取的常数列,W_2 为适当选取的核函数.计算(6.3.2)式中的 v_n 可用积分近似算法.虽然其计算量会随阶数 p 增加而加大,但是不同于"维数灾难"那样的困难.根据(6.3.1)式,适当地选取常数列 c_n,可使得在 H_1 成立时,

$$v_n\xrightarrow{p}\infty,\quad 当\ n\rightarrow\infty\ 时, \tag{6.3.3}$$

式中"$\xrightarrow{p}$"表示依概率收敛.不过,却无法选择 c_n 和 W_2,使得在 H_0 成立时给出 v_n 的渐近分布.因此,我们将从拟合优度的角度出发构造另一个统计量.

考虑平稳鞅差序列 $e_{p+1},\cdots,e_n$ 的如下加权和的平方,使得

$$(c_u^2/n)\left\{\sum_{t=p+1}^{n}e_t W_1(\boldsymbol{x}_{t-1})\right\}^2\xrightarrow{d}\chi_1^2,$$

其中 χ_1^2 是自由度为 1 的卡方随机变量,c_n^2 为适当选取的常数,W_1 为适当选取的核函数.上式意味着适当选取 c_u^2 和 W_1,可使随机变量$(c_u/\sqrt{n})e_t W_1(\boldsymbol{x}_{t-1})$之和的平方以 χ_1^2 为其渐近分布.为了得到检验统计量,将上式中的不可观测的 e_t 用(6.2.20)式中的 ξ_t 代替,则得到

$$u_n = \frac{C_u^2}{n}\left\{\sum_{t=p+1}^{n}\xi_t W_1(\boldsymbol{x}_{t-1})\right\}^2. \tag{6.3.4}$$

在 H_0 成立时，ξ_t 可近似为 e_t 的观测. 在不太强的条件下，上式中的 u_n 仍以 χ_1^2 为其渐近分布，即

$$u_n \xrightarrow{d} \chi_1^2. \tag{6.3.5}$$

将以上两类不同类型的统计量相加，即

$$\eta_n = u_n + v_n, \tag{6.3.6}$$

得到一个混合型的检验统计量. 由于 u_n 和 v_n 都是非负的，根据上面叙述的(6.3.3)式，当 H_1 成立时，有

$$\lim_{n\to\infty}\eta_n = \infty \quad (\text{依概率成立}). \tag{6.3.7}$$

在构造本节的混合型检验方法时，其中还有关键的一步是：在构造 v_n 时，要求当 H_0 成立时，有

$$\lim_{n\to\infty} v_n = 0 \quad (\text{依概率成立}).$$

此时联合(6.3.5)和(6.3.6)两式，当 H_0 成立时，有

$$\eta_n \xrightarrow{d} \chi_1^2. \tag{6.3.8}$$

由(6.3.7)和(6.3.8)式即知，η_n 是所要的混合型检验统计量.

混合型检验的具体检验步骤如下：

第一步：利用 K-S 检验方法的第一步，计算 LAR(p)模型拟合的残差 ξ_t 和残差方差 $\hat{\sigma}^2$.

第二步：计算以下诸量：

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t,$$

$$\hat{\gamma}_{i-j} = \frac{1}{n}\sum_{t=p+1}^{n}(x_{t-i}-\bar{x})(x_{t-j}-\bar{x}) \quad (1\leqslant i,j\leqslant p),$$

$$\overline{W}_1 = \frac{1}{n}\sum_{t=p+1}^{n} W_1(x_{t-1},\cdots,x_{t-p}),$$

其中

$$W_1(\boldsymbol{x}) = \left(2e^{-q\|\hat{\Gamma}^{-1/2}(\boldsymbol{x}-\bar{\boldsymbol{x}}_n)\|} - e^{-2q\|\hat{\Gamma}^{-1/2}(\boldsymbol{x}-\bar{\boldsymbol{x}}_n)\|}\right)\frac{q}{3},$$

$$\hat{\Gamma} = (\hat{\gamma}_{i-j})_{1\leqslant i,j\leqslant p}, \quad \bar{\boldsymbol{x}}_n = (\bar{x}_n,\cdots,\bar{x}_n)^{\tau}.$$

$$\hat{\sigma}_w^2 = \frac{1}{n}\sum_{t=p+1}^{n}\{W_1(x_{t-1},\cdots,x_{t-p}) - \overline{W}_1\}^2,$$

$$\hat{\gamma}_u = \frac{1}{n}\sum_{t=p+1}^{n} \boldsymbol{x}_{t-1}\{W_1(x_{t-1},\cdots,x_{t-p}) - \overline{W}_1\},$$

$$\hat{\sigma}_u^2 = \hat{\sigma}^2\{\hat{\sigma}_w^2 + \hat{\gamma}_u^{\tau}\hat{\Gamma}^{-1}\hat{\gamma}_u\},$$

$$u_n = \frac{1}{n\hat{\sigma}_u^2}\left\{\sum_{t=p+1}^{n}\hat{\xi}_t W_1(x_{t-1},\cdots,x_{t-p})\right\}^2.$$

第三步：适当选定 $q(\approx 1)$，记

$$W_2(\boldsymbol{s}) = \prod_{i=1}^{p}\left\{(2e^{-q|s_i|} - e^{-2q|s_i|})\frac{q}{3}\right\}.$$

计算

$$v_n = c_n\int\{(S_n(\boldsymbol{s}))\}^2 W_2(\boldsymbol{s})d\boldsymbol{s},$$

其中

$$c_n = n/\log\log n,\quad n \geqslant 27,$$

$$S_n(\boldsymbol{s}) = \frac{c_v}{n}\sum_{t=p+1}^{n}\hat{\xi}_t I(\boldsymbol{x}_{t-1} < \boldsymbol{s}),$$

$$c_v = \frac{1}{2\hat{\sigma}}.$$

注意，计算 v_n 时不必用积分计算，而用如下公式计算：

$$v_n = \frac{c_n c_v^2}{n^2}\sum_{k,j=p+1}^{n}\hat{\xi}_k\hat{\xi}_j\int I(\boldsymbol{x}_{k-1} < \mathbf{s})I(\boldsymbol{x}_{j-1} < \mathbf{s})W_2(\mathbf{s})d\mathbf{s}$$

$$= \frac{c_n c_v^2}{n}\sum_{k,j=p+1}^{n}\hat{\xi}_k\hat{\xi}_j\psi(\boldsymbol{x}_{k-1},\boldsymbol{x}_{j-1}),$$

其中

$$\begin{aligned}
&\psi(\boldsymbol{x}_{k-1},\boldsymbol{x}_{j-1}) \\
&= \int I(\boldsymbol{x}_{k-1} < \mathbf{s})I(\boldsymbol{x}_{j-1} < \mathbf{s})W_2(\mathbf{s})d\mathbf{s} \\
&= \prod_{i=1}^{p}\int I(x_{k-i} < s_i)I(x_{j-i} < s_i) \\
&\quad\times (2e^{-q|s_i|} - e^{-2q|s_i|})\frac{q}{3}ds_i \\
&= \prod_{i=1}^{p}\int_{-\infty}^{x_{k-i}\wedge x_{j-i}}(2e^{-q|s_i|} - e^{-2q|s_i|})\frac{q}{3}ds_i,
\end{aligned}$$

而且

$$\int_{-\infty}^{c}(2e^{-q|s|}-e^{-2q|s|})ds$$

$$=\begin{cases}2e^{qc}-\dfrac{1}{2}e^{2qc}, & \text{当 } c\leqslant 0;\\ 3-2e^{-qc}+\dfrac{1}{2}e^{-2qc}, & \text{当 } c>0.\end{cases}$$

第四步：利用文献[7]中已证明的(6.3.7)和(6.3.8)式，对置信水平 β，由 χ_1^2 分布可查出 χ_β 的值，使得

$$P(\chi_1^2>\chi_\beta)=\beta.$$

当 $\hat{\eta}_n=\hat{u}_n+\hat{v}_n>\chi_\beta$ 时，拒绝线性性假设 H_0.

§ 6.4 其他非参数型检验方法

一、双谱方法

考虑如下的平稳线性序列：

$$x_t=\sum_{k=0}^{\infty}\alpha_k\varepsilon_{t-k},\quad \sum_{k=0}^{\infty}|\alpha_k|<\infty, \tag{6.4.1}$$

其中$\{\varepsilon_t\}$为白噪声序列，且 $E\varepsilon_t=0$，$E\varepsilon_t^2=\sigma^2$，$E\varepsilon_t^6<\infty$. 由第 1 章的讨论知$\{x_t\}$的谱密度为

$$\begin{aligned}f_1(\omega)&=\frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma_k e^{-ik\omega}\\&=\frac{\sigma^2}{2\pi}\left|\sum_{j=0}^{\infty}\alpha_j e^{-ij\omega}\right|^2\\&=\frac{\sigma^2}{2\pi}|\alpha(\omega)|^2,\end{aligned} \tag{6.4.2}$$

其中

$$\gamma_k=Ex_tx_{t-k},$$

$$\alpha(\omega)=\sum_{j=0}^{\infty}\alpha_j e^{-ij\omega}. \tag{6.4.3}$$

$\{x_t\}$的三阶矩谱密度——即双谱密度为

$$\begin{aligned}f_2(\omega,\lambda)&=\frac{1}{(2\pi)^2}\sum_{k=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\gamma_{k,j}\ e^{-ik\omega-ij\lambda}\\&=\frac{c_2}{(2\pi)^2}\alpha(\omega)\alpha(\lambda)\alpha(-\omega-\lambda).\end{aligned} \tag{6.4.4}$$

由(6.4.2)～(6.4.4)式可知

$$\frac{|f_2(\omega,\lambda)|^2}{|f_1(\omega)f_1(\lambda)f_1(\omega+\lambda)|}$$
$$=\frac{c_2^2|\alpha(\omega)|^2|\alpha(\lambda)|^2|\alpha(\omega+\lambda)|^2}{2\pi\sigma^6|\alpha(\omega)|^2|\alpha(\lambda)|^2|\alpha(\omega+\lambda)|^2}$$
$$=\frac{c_2^2}{2\pi\sigma^6}. \tag{6.4.5}$$

在上式中,当 $E\varepsilon_t^3=0$ 时,$c_2=0$. 例如$\{\varepsilon_t\}$为正态白噪声序列时即是如此.

在非线性平稳序列情况下,仍可用(6.4.2)和(6.4.4)式的前半部分定义 $f_1(\omega)$和 $f_2(\omega,\lambda)$. 此时 $f_1(\omega)$和 $f_2(\omega,\lambda)$不再与(6.4.3)式的函数 α 有联系. 从而由(6.4.5)式定义的函数之比不再恒取常数. 依此事实,Subba Rao 和 Gabr(1980)(见文献[98])提出双谱检验方法,以检验平稳序列的线性性. 其检验步骤如下:

第一步:计算 γ_k 和 $\gamma_{k,j}$的样本值

$$\hat{\gamma}_k=\frac{1}{n}\sum_{t=k+1}^{n}(x_t-\bar{x}_n)(x_{t-k}-\bar{x}_n)$$
$$(k=0,1,\cdots,n-1),$$
$$\bar{x}_n=\frac{1}{n}\sum_{t=1}^{n}x_t,$$
$$\hat{\gamma}_{k,j}=\frac{1}{n}\sum_{t=p+1}^{n}(x_t-\bar{x}_n)(x_{t-k}-\bar{x}_n)(x_{t-j}-\bar{x}_n),$$
$$p=\max(k,j)\quad(k,j=1,2,\cdots,n-1).$$

第二步:计算

$$\hat{f}_1(\omega)=\frac{1}{2\pi}\sum_{k=-n+1}^{n-1}\hat{\gamma}_k e^{-ik\omega},$$
$$\hat{f}_2(\omega,\lambda)=\frac{1}{(2\pi)^2}\sum_{k=-n+1}^{n-1}\sum_{j=-n+1}^{n-1}\hat{\gamma}_{k,j}\ e^{-ik\omega-ij\lambda}.$$

注意,在上式中当 $\max(k,j)\geqslant n$ 时,取 $\hat{\gamma}_{k,j}=0$.

第三步:选用适当的谱窗方法(见文献[19]),将光滑的 $\hat{f}_1(\omega)$和 $\hat{f}_2(\omega,\lambda)$分别记为 $\tilde{f}_1(\omega)$,$\tilde{f}_2(\omega,\lambda)$.

第四步:计算如下统计量:

$$\hat{S}_n=\sum_{\omega_1}\sum_{\omega_2}\frac{|\tilde{f}_2(\omega_1,\omega_2)|^2}{|\tilde{f}_1(\omega_1)\tilde{f}_1(\omega_2)\tilde{f}_1(\omega_1+\omega_2)|},$$

其中双重求和的足标 ω_1 和 ω_2 分别取遍

$$\omega_1, \omega_2 = \frac{2\pi k}{n} \quad (k = 0, 1, \cdots, n).$$

当 $\{x_t\}$ 为线性序列时，$\hat{S}_n$ 渐近于 χ^2 分布. 于是根据给定的置信水平可求出否定域，这里从略.

上述检验方法，只涉及了平稳序列的三阶矩的结构特征，所以它不是相容的. 除此以外，上述方法对第三步中的谱窗的选择较敏感，在样本不太大时尤其如此. 虽然 Hinich（见文献[65]）对上述方法做了改进，但是仍然不具有相容性.

二、高阶矩检验

考虑平稳序列 $\{x_t\}$，用以下记号表示其自相关函数

$$\rho_k(x) = E(x_t - Ex_1)(x_{t-k} - Ex_1)/E(x_t - Ex_1)^2 \quad (6.4.6)$$

$$(k = 0, \pm 1, \cdots).$$

若 $\{x_t\}$ 满足某个平稳可逆的 LARMA(p,q) 模型，即

$$x_t = \varphi_1 x_{t-1} + \cdots + \varphi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}, \quad (6.4.7)$$

式中 $\{\varepsilon_t\}$ 为白噪声序列，满足(3.2.2)式，并且 $E\varepsilon_t^4 < \infty$. 于是，类似于(6.4.6)式的记法，$\rho_k(\varepsilon)$ 必满足

$$\rho_k(\varepsilon) = 0, \ k \neq 0. \quad (6.4.8)$$

由于 $\{\varepsilon_t^2\}$ 也是相互独立同分布的，从而有

$$\rho_k(\varepsilon^2) = 0, \ k \neq 0. \quad (6.4.9)$$

根据上述事实，Maravall、Mcleod 和 Li（见文献[76]和[79]）引进了如下的检验线性性的方法，其步骤如下：

第一步：根据样本 $x_1, \cdots, x_n$，拟合 LARMA(p,q) 模型(6.4.7). 具体方法可见文献[20]、[125]、[133]中所述. 得到拟合残差 $\hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_n$，及平方和的值，即

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n} \hat{\varepsilon}_t^2.$$

第二步：计算

$$\hat{\rho}_k(\hat{\varepsilon}^2) = \sum_{t=k+1}^{n} (\hat{\varepsilon}_t^2 - \hat{\sigma}^2)(\hat{\varepsilon}_{t-k}^2 - \hat{\sigma}^2) \Big/ \sum_{t=1}^{n} (\hat{\varepsilon}_t^2 - \hat{\sigma}^2)^2$$

$$(k = 1, \cdots, m).$$

第三步：计算

$$Q_n = \frac{n(n-2)}{n-k} \sum_{k=1}^{m} \hat{\rho}_k^2(\hat{\varepsilon}^2).$$

当$\{x_t\}$为线性序列时,上述Q_n的分布渐近于χ^2_m,其中m为适当选取的正整数.根据这一事实,对给定的置信水平,可给出否定域,这里从略.

以上方法是高阶矩方法的一种.在实际使用中,阶数p和q是未知的,可用AIC准则及实际数据进行估计.因此,其第一步可借助于熟知的LARMA建模方法完成.第三步中的m是比n小得多的正整数,可随n增大而适当大些.需要说明的是:在$\{\varepsilon_t\}$为平稳鞅差序列时,这里所述的方法就不适用了.

三、Tukey 方法

考虑LAR模型的检验问题,即检验$\{x_t\}$是否满足如下模型:

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + \varepsilon_t, \tag{6.4.10}$$

其中$\{\varepsilon_t\}$为白噪声序列,系数$\boldsymbol{\alpha}=(\alpha_1,\cdots,\alpha_p)^\tau$满足(1.2.5)式.Keenan在1985年(见文献[69])提出如下检验方法:

第一步:用K-S检验法中的第一步,对数据$x_1,\cdots,x_n$拟合LAR(p)模型(6.4.10),得到拟合残差$\hat{\varepsilon}_{p+1},\cdots,\hat{\varepsilon}_n$,及其平方和

$$\mathrm{RSS}=\sum_{t=p+1}^{n}\hat{\varepsilon}_t^2,$$

$$\hat{x}_t=\hat{\alpha}_0+\hat{\alpha}_1 x_{t-1}+\cdots+\hat{\alpha}_p x_{t-p}\quad (t=p+1,\cdots,n).$$

第二步:记$y_t=\hat{x}_t^2$,并作y_t与$(1,x_{t-1},\cdots,x_{t-p})$的线性回归拟合,得到拟合残差$\hat{\delta}_{p+1},\cdots,\hat{\delta}_n$.

第三步:对于二元序列$(\hat{\varepsilon}_t,\hat{\delta}_t)$ $(t=p+1,\cdots,n)$,进行一元线性回归拟合,即拟合如下回归模型:

$$\hat{\varepsilon}_t=\beta\hat{\delta}_t+\eta_t.$$

计算出β的最小二乘估计$\hat{\beta}$,以及

$$b=\hat{\beta}\Big(\sum_{t=p+1}^{n}\hat{\delta}_t^2\Big)^{1/2},$$

$$F_n=\frac{b^2(n-2p-2)}{\mathrm{RSS}-b^2}.$$

在$\{x_t\}$满足(6.4.10)式时,F_n的分布近似为$F_{1,n-2p-2}$(见文献[106]).依此不难给出以后的检验步骤,这里从略.

§ 6.5　Lagrange 乘子检验方法

本章§6.2～§6.4介绍的检验方法,无论是针对预报线性性检验,

还是针对结构线性性检验，都属于非参数型检验方法. 在这一节，我们介绍一种参数型检验方法——Lagrange 乘子检验方法，简称 LM 方法，借以解决如下的对立假设检验问题：

H_0：平稳序列$\{x_t\}$满足如下线性模型：

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + \varepsilon_t, \tag{6.5.1}$$

式中$\{\varepsilon_t\}$为白噪声序列，或为平稳鞅差序列，满足(3.2.2)式；

H_1：$\{x_t\}$为某指定的有限参数型非线性模型.

例如，在对立假设 H_1 中，当平稳序列$\{x_t\}$指定为如下的双线性模型：

$$x_t = \sum_{j=1}^{p} \alpha_j x_{t-j} + \sum_{i=1}^{P}\sum_{j=1}^{Q} \beta_{ij} x_{t-j} \varepsilon_{t-j} + \varepsilon_t \tag{6.5.2}$$

时，参数型的对应假设为：

$$H_0: \beta_{ij} = 0 \quad (i=1,\cdots,P; j=1,\cdots,Q); \tag{6.5.3}$$

$$H_1: \sum_{i,j} |\beta_{ij}| \neq 0. \tag{6.5.4}$$

又如，当 H_1 中平稳序列指定为 ARCH 模型时，即

$$\begin{cases} x_t = \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + e_t; \\ e_t = \varepsilon_t (\varphi_0 + \varphi_1 e_{t-1}^2 + \cdots + \varphi_p e_{t-p}^2)^{1/2}, \end{cases} \tag{6.5.5}$$

式中$\{\varepsilon_t\}$为正态白噪声序列，$\varepsilon_t \sim N(0,1)$，$\varphi_0 > 0$，$\varphi_i \geqslant 0 \quad (i=1,\cdots,p)$，$\varphi_1 + \cdots + \varphi_p < 1$，此时，参数型的对应假设为：

$$H_0: \varphi_i = 0 \quad (i=1,\cdots,p); \tag{6.5.6}$$

$$H_1: \sum_{i=1}^{p} \varphi_i \neq 0. \tag{6.5.7}$$

在统计分析中，对参数进行假设检验时，构造检验统计量的方法通常有三种：一种是以参数估计的渐近正态性为基础的 Wold 方法（简称 W 方法），例如常见的 t-检验，F-检验等都属于 W 方法. 另一种是似然比方法（简称 LR 方法），也是我们比较熟悉而经常使用的方法. 第三种方法就是本节将要介绍的 Lagrange 乘子方法（简称 LM 方法），它是将零假设条件看成一个约束条件，通过对有约束的极大似然函数的一阶偏导数进行检验，对参数假设作出判断. 在某些模型检验中，LM 方法不仅能带来计算上的方便，而且所构造的检验统计量与上述两种方法所构成的检验统计量具有相同的渐近性质.

设 $x_1,\cdots,x_n$ 是来自某个分布族的样本. 该分布族中的分布除参数 $\boldsymbol{\theta} = (\theta_1,\cdots,\theta_p)^{\tau}$之外都是已知的. 记 $L(\boldsymbol{\theta})$为样本$(x_1,\cdots,x_n)$的对数似然函数. 考虑如下关于 $\boldsymbol{\theta}$ 的假设检验问题：

$$H_0: h_j(\theta)=0 \quad (j=1,\cdots,k;\ k<p). \tag{6.5.8}$$

将上述零假设中的 k 个方程视为关于参数 θ 的 k 个约束条件，利用 Lagrange乘子法，求如下的有约束条件的极大似然估计，即求函数

$$f(\theta,\lambda_1,\cdots,\lambda_k) = L(\theta) + \sum_{j=1}^{k}\lambda_j h_j(\theta) \tag{6.5.9}$$

关于 $\theta,\lambda_1,\cdots,\lambda_k$ 的极大值. 对于(6.5.9)式，关于 θ 和 $\boldsymbol{\lambda}=(\lambda_1,\cdots,\lambda_k)^{\tau}$，求一阶偏导数，并令其为零，得到

$$\frac{\partial f}{\partial \theta_i} = \frac{\partial L(\theta)}{\partial \theta_i} + \sum_{j=1}^{k}\lambda_j \frac{\partial h_j(\theta)}{\partial \theta_i} = 0 \quad (i=1,\cdots,p), \tag{6.5.10}$$

$$\frac{\partial f}{\partial \lambda_i} = h_i(\theta) = 0 \quad (i=1,\cdots,k). \tag{6.5.11}$$

令

$$\boldsymbol{D}(\theta) = \begin{bmatrix} \dfrac{\partial L(\theta)}{\partial \theta_1} \\ \cdots \\ \dfrac{\partial L(\theta)}{\partial \theta_p} \end{bmatrix},$$

$$H(\theta) = \begin{bmatrix} \dfrac{\partial h_1(\theta)}{\partial \theta_1} & \cdots & \dfrac{\partial h_1(\theta)}{\partial \theta_p} \\ \cdots & \cdots & \cdots \\ \dfrac{\partial h_k(\theta)}{\partial \theta_1} & \cdots & \dfrac{\partial h_k(\theta)}{\partial \theta_p} \end{bmatrix},$$

则(6.5.10)式和(6.5.11)式可写成如下矩阵方程：

$$\boldsymbol{D}(\theta) + H(\theta)\boldsymbol{\lambda} = \mathbf{0}. \tag{6.5.12}$$

若零假设 H_0 为真，则 θ 的有约束极大似然估计 $\hat{\theta}_{H_0}$ 近似于 θ 的无约束极大似然估计 $\hat{\theta}_{ML}$，从而 $\boldsymbol{D}(\hat{\theta}_{H_0})$ 将近似于零向量. 在某些正则条件下，Feigin 在 1976 年(见文献[49])证明了：在 H_0 成立时，$\boldsymbol{D}(\hat{\theta}_{H_0})$ 是一个零均值鞅，且在 Growder(见文献[40])给出的条件下，

$$C_n^{-1/2}\boldsymbol{D}(\hat{\theta}_{H_0}) \xrightarrow{d} N(0, W(\theta)),$$

式中

$$W(\theta) = \lim_{n\to\infty} C_n^{-1/2} I(\theta) C_n^{-1/2},$$

其中 $I(\theta)=E(-\partial^2 L(\theta)/\partial\theta\partial\theta^{\tau})$ 是 Fisher 信息矩阵，C_n 是适当选取的矩阵序列. 根据以上结果，可以构造如下 LM 检验统计量：

$$\mathrm{LM} = \boldsymbol{D}(\hat{\boldsymbol{\theta}}_{H_0})^{\tau}\boldsymbol{I}(\hat{\boldsymbol{\theta}}_{H_0})^{-1}\boldsymbol{D}(\hat{\boldsymbol{\theta}}_{H_0}). \tag{6.5.13}$$

可以证明在通常的极大似然估计渐近理论的正则条件下，LM 检验与 W 检验和 LR 检验是渐近等价的. 也就是说，在 H_0 成立时，

$$\mathrm{LM} \xrightarrow{d} \chi_k^2, \quad (\text{当 } n \to \infty). \tag{6.5.14}$$

从而对给定的置信水平 β，不难确定 H_0 的否定域，这里从略.

在存在多余参数的情形，记其为 $\nu=(\nu_1,\cdots,\nu_l)^{\tau}$，我们需要修正表达式(6.5.13)式. 令 $\tilde{\boldsymbol{\theta}}=(\boldsymbol{\theta}^{\tau},\boldsymbol{\nu}^{\tau})^{\tau}$，考虑如下假设检验问题：

$H_0: h_j(\tilde{\boldsymbol{\theta}})=0 \quad (j=1,\cdots,k).$

按前面所述的方法求得方程(6.5.12)和 $I(\tilde{\boldsymbol{\theta}})$. 根据 $\boldsymbol{\theta}$ 和 $\boldsymbol{\nu}$ 的维数把 $\boldsymbol{D}(\tilde{\boldsymbol{\theta}})$ 和 $I(\tilde{\boldsymbol{\theta}})$ 写成分块向量或分块矩阵，即

$$\boldsymbol{D}(\tilde{\boldsymbol{\theta}})=\begin{pmatrix}\boldsymbol{D}_1(\tilde{\boldsymbol{\theta}})\\ \boldsymbol{D}_2(\tilde{\boldsymbol{\theta}})\end{pmatrix}, \quad I(\tilde{\boldsymbol{\theta}})=\begin{pmatrix}I_{11}(\tilde{\boldsymbol{\theta}}) & I_{12}(\tilde{\boldsymbol{\theta}})\\ I_{21}(\tilde{\boldsymbol{\theta}}) & I_{22}(\tilde{\boldsymbol{\theta}})\end{pmatrix}.$$

由(6.5.10)式和(6.5.12)式可知 $\boldsymbol{D}_2(\tilde{\boldsymbol{\theta}})=0$. 从而由(6.5.13)式和分块矩阵求逆公式(见文献[129])，对零假设 H_0，LM 检验统计量为

$$\mathrm{LM}_1 = \boldsymbol{D}_1(\tilde{\boldsymbol{\theta}}_{H_0})^{\tau}[I_{11}(\tilde{\boldsymbol{\theta}}) - I_{12}(\tilde{\boldsymbol{\theta}})I_{22}^{-1}(\tilde{\boldsymbol{\theta}})I_{21}(\tilde{\boldsymbol{\theta}})]^{-1}\boldsymbol{D}_1(\tilde{\boldsymbol{\theta}}_{H_0}), \tag{6.5.15}$$

其中 $\tilde{\boldsymbol{\theta}}_{H_0}$ 为参数 $\tilde{\boldsymbol{\theta}}$ 的有约束极大似然估计. 同样，在一些正则条件下，对于由(6.5.15)式给出的检验统计量 LM_1，(6.5.14)式仍成立.

现在讨论(6.5.3)式和(6.5.4)式以及(6.5.6)式和(6.5.7)式提出的假设检验问题.

设 $x_1,\cdots,x_n$ 是来自某平稳序列的样本. 我们要检验 $x_1,\cdots,x_n$ 是否为来自双线性模型(6.5.2)的样本，即 $x_1,\cdots,x_n$ 是否满足如下非线性模型：

$$x_t = \sum_{j=1}^{p}\alpha_j x_{t-j} + \sum_{i=1}^{P}\sum_{j=1}^{Q}\beta_{ij}x_{t-i}\varepsilon_{t-j} + \varepsilon_t. \tag{6.5.16}$$

这时(6.5.8)式的零假设等价于

$$H_{\mathrm{BL0}}: \beta_{ij} = 0 \quad (i = 1,\cdots,P;\ j = 1,\cdots,Q); \tag{6.5.17}$$

相应的对立假设为

$$H_{\mathrm{BL}}: \sum_{i,j}|\beta_{ij}| \neq 0. \tag{6.5.18}$$

在 H_{BL} 下，$(x_1,\cdots,x_n)$ 的对数似然函数为

$$L(\tilde{\boldsymbol{\theta}}) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2}\sigma^{-2}\sum_{t=1}^{n}\varepsilon_t^2, \tag{6.5.19}$$

其中 $\tilde{\boldsymbol{\theta}}=(\boldsymbol{\theta}^{\tau},\ \boldsymbol{\nu}^{\tau})^{\tau}$，$\nu=(\alpha_1,\cdots,\alpha_p)^{\tau}$，$\boldsymbol{\theta}=(\beta_{11},\cdots,\beta_{1p},\cdots,\beta_{s1},\cdots,\beta_{sp})^{\tau}$. 由

(6.5.16)式不难算得

$$\frac{\partial L(\theta)}{\partial \beta_{ij}} = -x_{t-i}\varepsilon_{t-j} \quad (i=1,\cdots,P;\ j=1,\cdots,Q), \tag{6.5.20}$$

$$\frac{\partial L(\theta)}{\partial \alpha_i} = -x_{t-i} \quad (i=1,\cdots,p). \tag{6.5.21}$$

经过繁琐的代数运算，对 H_{BL0} 而言，LM 检验统计量为

$$\mathrm{LM}(BL) = \hat{\sigma}^{-2}\left(\sum_{t=1}^{n} \mathbf{z}_{1t}\hat{\varepsilon}_t\right)^{\tau}(M_{11} - M_{10}M_{00}^{-1}M_{01})^{-1}\left(\sum_{t=1}^{n} \mathbf{z}_{1t}\hat{\varepsilon}_t\right), \tag{6.5.22}$$

式中

$$M_{0i} = M_{i0}^{\tau} = \sum_{t=1}^{n} \mathbf{z}_{it}\mathbf{z}_{0t}^{\tau} \quad (i=0,1),$$

$$M_{11} = \sum_{t=1}^{n} \mathbf{z}_{1t}\mathbf{z}_{1t}^{\tau},$$

$$\mathbf{z}_{0t} = (x_{t-1},\cdots,x_{t-p})^{\tau},$$

$$\mathbf{z}_{1t} = (\hat{\varepsilon}_{t-1}x_{t-1},\hat{\varepsilon}_{t-1}x_{t-2},\cdots,\hat{\varepsilon}_{t-Q}x_{t-P})^{\tau},$$

$$\hat{\varepsilon}_t = x_t - \hat{\alpha}_1 x_{t-1} - \cdots - \hat{\alpha}_p x_{t-p} \quad (t=1,\cdots,n),$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_t^2,$$

其中 $\hat{\alpha}_1,\cdots,\hat{\alpha}_p$ 是在 H_{BL0} 成立的条件下，$\alpha_1,\cdots,\alpha_p$ 的极大似然估计. 利用 LM 检验的统计分析，在 H_0 下，有

$$\mathrm{LM}(BL) \xrightarrow{d} \chi^2_{PQ}.$$

这样，对给定的置信水平 β，容易确定 H_{BL0} 的否定域，在此从略.

下面讨论 ARCH 模型的检验. 考虑如下 ARCH 模型：

$$\begin{cases} x_t = \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + e_t; \\ e_t \mid \mathscr{F}_{t-1}^{x} \sim N(0,h_t); \\ h_t = \varphi_0 + \varphi_1 e_{t-1}^2 + \cdots + \varphi_q e_{t-q}^2. \end{cases} \tag{6.5.23}$$

设 $x_1,\cdots,x_n$ 是某平稳序列的一组样本. 我们要检验这组样本是否来自模型(6.5.23). 等价地，可考虑假设检验问题(6.5.6)和(6.5.7)，即零假设为

$$H_{\mathrm{ARCH0}}:\ \varphi_i = 0 \quad (i=1,\cdots,q);$$

而对立假设为

$$H_{\text{ARCH}}:\quad \sum_{i=1}^{q}\varphi_i \neq 0.$$

在假设 H_{ARCH} 下，$(x_1,\cdots,x_n)$ 关于 $(x_0,x_{-1},\cdots,x_{-p+1})$ 的条件对数似然函数为

$$L(\tilde{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\log f(x_t|x_{t-1},\cdots,x_{t-p}), \qquad (6.5.24)$$

式中 $\tilde{\theta}=(\theta^{\tau},\nu^{\tau})^{\tau}$，$\theta=(\varphi_0,\varphi_1,\cdots,\varphi_q)^{\tau}$，$\nu=(\alpha_1,\cdots,\alpha_p)^{\tau}$，

$$f(x_t|x_{t-1},\cdots,x_{t-p}) = (2\pi h_t)^{-1/2}\exp\{-(2h_t)^{-1}e_t^2\}.$$

忽略常数项后，并记 $l_t=-\frac{1}{2}\log h_t-(2h_t)^{-1}e_t^2$，则(6.5.24)式可近似地表示为

$$\begin{aligned} L(\tilde{\theta}) &= \frac{1}{n}\sum_{t=1}^{n}l_t \\ &= -\frac{1}{n}\sum_{t=1}^{n}\log h_t - \frac{1}{2n}\sum_{t=1}^{n}\frac{e_t^2}{h_t}. \end{aligned} \qquad (6.5.25)$$

对(6.5.25)式关于 θ 求一阶和二阶偏导数，以及关于 θ 和 ν 求混合二阶偏导数，可得

$$\begin{aligned} \frac{\partial l_t}{\partial \theta} &= -\frac{1}{2h_t}\frac{\partial h_t}{\partial \theta} + \frac{e_t^2}{2h_t^2}\frac{\partial h_t}{\partial \theta} \\ &= \frac{1}{2h_t}\frac{\partial h_t}{\partial \theta}\left(\frac{e_t^2}{h_t}-1\right), \end{aligned} \qquad (6.5.26)$$

$$\begin{aligned} \frac{\partial^2 l_t}{\partial \theta\partial \theta^{\tau}} &= \left(\frac{e_t^2}{h_t}-1\right)\frac{\partial}{\partial \theta^{\tau}}\left(\frac{1}{2h_t}\frac{\partial h_t}{\partial \theta}\right) \\ &\quad - \frac{e_t^2}{2h_t^3}\cdot\frac{\partial h_t}{\partial \theta}\cdot\frac{\partial h_t}{\partial \theta^{\tau}}, \\ \frac{\partial^2 l_t}{\partial \theta\partial \nu^{\tau}} &= \left(\frac{e_t^2}{h_t}-1\right)\frac{\partial}{\partial \nu^{\tau}}\left(\frac{1}{2h_t}\frac{\partial h_t}{\partial \theta}\right) \\ &\quad + \frac{1}{2h_t}\cdot\frac{\partial h_t}{\partial \theta}\left(\frac{2e_t}{h_t}\frac{\partial e_t}{\partial \nu^{\tau}} - \frac{e_t^2}{h_t^2}\frac{\partial h_t}{\partial \nu^{\tau}}\right). \end{aligned}$$

注意到 $E\{e_t|\mathscr{F}_{t-1}^x\}=0$，$E\{e_t^2|\mathscr{F}_{t-1}^x\}=h_t$，

$$\begin{aligned} \frac{\partial h_t}{\partial \alpha_i} &= -\frac{\partial}{\partial \alpha_i}\left[\varphi_0 + \sum_{j=1}^{q}\varphi_j\left(x_{t-j} - \sum_{k=1}^{p}\alpha_j x_{t-j-k}\right)^2\right] \\ &= -2\sum_{j=1}^{q}\varphi_j e_{t-j}x_{t-j-i}, \end{aligned}$$

由条件期望的性质，h_t 关于 $\mathscr{F}_{t-1}^x$ 可测及 $E\{e_t^2|\mathscr{F}_{t-1}^x\}=h_t$，有

$$E\left(-\frac{\partial^2 l_t}{\partial \theta \partial \theta^\tau}\right) = E\left(\frac{1}{2h_t^2}\frac{\partial h_t}{\partial \theta}\frac{\partial h_t}{\partial \theta^\tau}\right),$$

$$E\left(-\frac{\partial^2 lt}{\partial \theta \partial \nu^\tau}\right) = 0.$$

记
$$z_t = (1, e_{t-1}^2, \cdots, e_{t-p}^2)^\tau,$$
$$f_t = (e_t^2/h_t) - 1 \quad (t = 1, \cdots, n),$$

则
$$h_t = z_t^\tau \theta.$$

从而有

$$\frac{\partial^2 h_t}{\partial \theta} = z_t.$$

于是,由(6.5.25)式和(6.2.26)式,有

$$D(\theta) = \frac{\partial L(\theta)}{\partial \theta} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t}{\partial \theta}.$$

$$= \frac{1}{n}\sum_{t=1}^{n}\frac{1}{2h_t}\frac{\partial h_t}{\partial \theta}\left(\frac{e_t^2}{h_t} - 1\right)$$

$$= \frac{1}{n}\sum_{t=1}^{n}\frac{1}{2h_t}z_t f_t, \tag{6.5.27}$$

$$I_{11}(\theta) = E\left(-\left(\frac{\partial^2 L(\theta)}{\partial \theta \partial \theta^\tau}\right)\right)$$

$$= \frac{1}{n}\sum_{t=1}^{n}E\left(\frac{1}{2h_t^2}\frac{\partial h_t}{\partial \theta}\frac{\partial h_t}{\partial \theta^\tau}\right)$$

$$= \frac{1}{n}\sum_{t=1}^{n}E\left(\frac{1}{2h_t^2}z_t z_t^\tau\right). \tag{6.5.28}$$

这样,可按如下步骤构造对假设 H_{ARCH0}的 LM 检验:

第一步:在 H_{ARCH0}下,求参数 ν 的极大似然估计 $\hat{\nu}$,即求 LAR(p)模型(6.5.1)参数的极大似然估计,并定义

$$\hat{e}_t = x_t - \hat{\alpha}_1 x_{t-1} - \cdots - \hat{\alpha}_p x_{t-p} \quad (t = 1, 2, \cdots, n), \tag{6.5.29}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}\hat{e}_t^2. \tag{6.5.30}$$

第二步:在 H_{ARCH0}下,注意到 h_t 为常数,记为 σ^2,其极大似然估计由(6.5.30)式定义. 令

$$\hat{z}_t = (1, \hat{e}_{t-1}^2, \cdots, \hat{e}_{t-q}^2)^\tau,$$
$$\hat{f}_t = (\hat{e}_t^2/\hat{\sigma}^2) - 1 \quad (t = 1, \cdots, n),$$
$$\hat{Z}_n = (\hat{z}_1^\tau, \cdots, \hat{z}_n^\tau)^\tau, \quad \hat{F}_n = (\hat{f}_1, \cdots, \hat{f}_n)^\tau,$$

则由(6.2.27)和(6.2.28)两式有

$$\boldsymbol{D}(\tilde{\boldsymbol{\theta}}_{H_0}) = \frac{1}{n}\sum_{t=1}^{n}\frac{1}{2\hat{\sigma}^2}\hat{z}_t\hat{f}_t = \frac{1}{2n\hat{\sigma}^2}\hat{Z}_n^{\tau}\hat{F}_n, \tag{6.5.31}$$

$$I_{11}(\tilde{\boldsymbol{\theta}}_{H_0}) = \frac{1}{n}\sum_{t=1}^{n}\frac{1}{2\hat{\sigma}^4}\hat{z}_t\hat{z}_t^{\tau} = \frac{1}{2n\hat{\sigma}^4}\hat{Z}_n^{\tau}\hat{Z}_n, \tag{6.5.32}$$

其中 $\tilde{\boldsymbol{\theta}}_{H0}=(\boldsymbol{\theta}^{\tau},\hat{\boldsymbol{\nu}}^{\tau})^{\tau}$.

第三步:由(6.5.31)和(6.5.32)式,可定义 LM 检验统计量为

$$\begin{aligned}\mathrm{LM(ARCH)} &= \boldsymbol{D}(\tilde{\boldsymbol{\theta}}_{H_0})^{\tau}I_{11}(\tilde{\boldsymbol{\theta}}_{H_0})^{-1}\boldsymbol{D}(\tilde{\boldsymbol{\theta}}_{H_0}) \\ &= \frac{1}{2n}\hat{F}_n^{\tau}\hat{Z}_n(\hat{Z}_n^{\tau}\hat{Z}_n)^{-1}\hat{Z}_n^{\tau}\hat{F}_n.\end{aligned} \tag{6.5.33}$$

利用 LM 检验的统计分析,在 H_{ARCH0} 下,

$$\mathrm{LM(ARCH)} \xrightarrow{d} \chi_q^2, \quad 当\ n\to\infty. \tag{6.5.34}$$

于是,对给定的置信水平 β,不难确定零假设 H_{ARCH0}的否定域.在此从略.

第 7 章

模型的估计

在这一章中，我们讨论非线性时间序列的估计问题，也称为非线性时间序列建模问题. 所谓时间序列建模问题，就是根据观测到的数据序列，估计出一个适当的模型去近似地描述它. 这也称为模型的识别和拟合问题. 对于非线性时间序列模型而言，模型识别包含两方面的内容：一方面是判断该数据序列是否有非线性性，也即对数据序列进行线性性检验，这已在第 6 章中讨论过了. 另一方面是当描述所观测到的数据序列的模型的类型选定后，确定模型的阶数，也称为模型定阶问题. 对线性序列情况，已有许多非常有效的方法，例如，AIC 准则、BIC 准则等（见文献[8]、[59]和[3]等). 而对非线性情况，这方面的方法还很少，我们将在 § 7.3 中作简单介绍. 在确定了模型的类型和阶数之后，就是如何估计模型的问题或称为模型拟合问题. 正如第 3 章所述，有参数型的非线性时间序列模型，也有非参数型的非线性时间序列模型. 因此，模型估计也有参数估计方法与非参数估计方法，这些内容将分别在 § 7.1 和 § 7.2 中进行讨论.

§ 7.1 参数型非线性模型的估计

在线性时间序列模型估计中，例如 LARMA 模型，常用的估计方法有最小二乘估计（简记为 LSE），矩估计（简记为 ME），以及极大似然估计（简记为 MLE). 由于 LSE 和 MLE 是在某种优化准则下求解的，故称为精细估计. ME 是利用矩与参数的关系，并用样本矩代替后求解的，一般精度较 LSE 和 MLE 差，常将其作为某些精细估计迭代求解的初始值. 除了上述三种估计方法外，还有条件最小二乘估计（简记为 CLSE）和条件极大似然估计（简记为 CMLE). 本节主要讨论参数型非线性时间序列模型参数的 CLSE、ME 和 CMLE. 在本节及以后两节中，我们讨论的模型都是非线性结构模型，并总假定这些模型满足第 4 章中所要求的平稳条件.

一、条件最小二乘估计

设$\{x_1,\cdots,x_n\}$是来自时间序列$\{x_n,n\geqslant 1\}$的观测序列，$\{x_n\}$的概率分布依赖于某未知参数$\theta=(\theta_1,\cdots,\theta_p)^\tau$，其中$\theta\in\Theta$，$\Theta$是$R^p$中的一个开集. 记$\mathscr{F}^x_{1,k}=\sigma(x_s,1\leqslant s\leqslant k)$为由$\{x_s,1\leqslant s\leqslant k\}$所生成的$\sigma$-代数，$\mathscr{F}_0$为平凡$\sigma$-代数. 我们可通过关于$\theta$极小化如下的残差平方和

$$Q_n(\theta)=\sum_{k=2}^{n}\{x_k-E_\theta\{x_k|\mathscr{F}^x_{1,k-1}\}\}^2 \tag{7.1.1}$$

来获得θ的估计，记其为$\hat{\theta}_n$. 这里记号E_θ表示x_t关于以θ为其参数的概率分布求均值. 此时，$\hat{\theta}_n$应满足

$$Q_n(\hat{\theta}_n)=\inf_{\theta\in\Theta}Q_n(\theta). \tag{7.1.2}$$

具体求解时，$\hat{\theta}_n$可通过如下方程得到

$$\frac{\partial}{\partial\theta_i}Q_n(\theta_n)=0\quad(i=1,\cdots,p). \tag{7.1.3}$$

这里总假定(7.1.3)式中的偏导数是存在的. 如果在真值$\theta^0=(\theta_1^0,\cdots,\theta_p^0)^\tau$的某邻域$S$中$E_\theta\{x_k|\mathscr{F}^x_{1,k-1}\}$是关于$\theta$二次连续可微的，则对$\delta>0$，$\|\theta-\theta^0\|<\delta$，且$S_{\theta^0}=\{\theta:\ \|\theta-\theta^0\|<\delta\}\subset S$，由 Taylor 展开式有

$$\begin{aligned}Q_n(\theta)&=Q_n(\theta^0)+(\theta-\theta^0)^\tau\left(\frac{\partial Q_n(\theta)}{\partial\theta}\right)_{\theta=\theta^0}\\&\quad+\frac{1}{2}(\theta-\theta^0)^\tau\left(\frac{\partial^2Q_n(\theta)}{\partial\theta\partial\theta^\tau}\right)_{\theta=\theta^*}(\theta-\theta^0)\\&=Q_n(\theta^0)+(\theta-\theta^0)^\tau\left(\frac{\partial Q_n(\theta)}{\partial\theta}\right)\\&\quad+\frac{1}{2}(\theta-\theta^0)^\tau V_n(\theta-\theta^0)\\&\quad+\frac{1}{2}(\theta-\theta^0)^\tau T_n(\theta^*)(\theta-\theta^0),\end{aligned} \tag{7.1.4}$$

其中θ^*是S_{θ^0}的适当的内点，V_n是由$Q_n(\theta)$的二阶偏导数在θ^0处的值所构成的$p\times p$矩阵，且

$$T_n(\theta^*)=\left(\frac{\partial^2Q_n(\theta)}{\partial\theta\partial\theta^\tau}\right)_{\theta=\theta^*}-V_n. \tag{7.1.5}$$

不难验证

$$\left(\frac{1}{2n}V_n\right)_{ij}=\sum_{k=2}^{n}\left[\frac{1}{n}\left(\frac{\partial}{\partial\theta_i}E_\theta\{x_k|\mathscr{F}^x_{1,k-1}\}\right)\right.$$

$$\times \left(\frac{\partial}{\partial \theta_j} E_\theta \{ x_k | \mathscr{F}^x_{1,k-1} \} \right)$$

$$- \frac{1}{n} \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} E_\theta \{ x_k | \mathscr{F}^x_{1,k-1} \} \right)$$

$$\times \left(x_k - E_{\theta^0} \{ x_k | \mathscr{F}^x_{1,k-1} \} \right) \Big]_{\theta = \theta^0}, \tag{7.1.6}$$

其中所有偏导数都在 θ^0 处取值. 对任意的 $1 \leqslant i, j \leqslant p$, 由鞅差强大数律, 可以验证, 当 $n \to \infty$ 时, 有

$$\frac{1}{n} \sum_{k=2}^{n} \left\{ \frac{\partial^2}{\partial \theta_i \partial \theta_j} E_\theta \{ x_k | \mathscr{F}^x_{1,k-1} \} \right\}_{\theta = \theta^0}$$

$$\times \{ x_k - E_{\theta^0} \{ x_k | \mathscr{F}^x_{1,k-1} \} \} \to 0, \text{ a.s.}, \tag{7.1.7}$$

且由遍历性, 也不难验证, 当 $n \to \infty$ 时, 有

$$\left\{ \frac{1}{n} \sum_{k=2}^{n} \left\{ \frac{\partial}{\partial \theta_i} E_\theta \{ x_k | \mathscr{F}^x_{1,k} \} \right\}_{\theta = \theta^0} \right.$$

$$\left. \times \left\{ \frac{\partial}{\partial \theta_j} E_\theta \{ x_k | \mathscr{F}^x_{1,k-1} \} \right\}_{\theta = \theta^0} \right\}_{1 \leqslant i,j \leqslant p} \to V, \text{ a.s.}, \tag{7.1.8}$$

其中 V 是一 $p \times p$ 的正定对称矩阵. 如果在适当的意义下能假定 T_n 是很小的, 则由(7.1.4)、(7.1.7)和(7.1.8)三式可知, $\hat{\theta}_n$ 是 θ^0 的强相容估计. 上述的想法可叙述为如下定理, 其详细证明见文献[56]和[71], 或[106].

定理 7.1.1　假定

$$\lim_{n \to \infty} \sup_{\delta \downarrow 0} (n\delta)^{-1} | (T_n(\theta^*))_{ij} | < \infty, \text{ a.s.}, \tag{7.1.9}$$

$$(1 \leqslant i, j \leqslant p),$$

$$(2n)^{-1} V_n \to V, \text{a.s.}, \tag{7.1.10}$$

其中 V 是 $p \times p$ 的常数正定对称矩阵, 且

$$n^{-1} \left(\frac{\partial}{\partial \theta_i} Q_n(\theta) \right)_{\theta = \theta^0} \to 0, \text{ a.s.} \quad (1 \leqslant i \leqslant p), \tag{7.1.11}$$

则存在一估计序列 $\{\hat{\theta}_n\}$, 使得 $\hat{\theta}_n \to \theta^0$, a.s., 并对任意的 $\varepsilon > 0$, 存在一事件 G, 满足 $P(G) > 1 - \varepsilon$ 和一个常数 n_0, 使得在 G 上, 当 $n > n_0$ 时, $\hat{\theta}_n$ 满足最小二乘方程(7.1.3), 且 $Q_n(\theta)$ 在 $\hat{\theta}_n$ 达到最小值.

下面讨论定理 7.1.1 中 $\hat{\theta}_n$ 的渐近正态性. 在 θ^0 处展开 $n^{-1/2} \frac{\partial}{\partial \theta} Q_n(\hat{\theta}_n)$, 得

$$0 = n^{-1/2} \frac{\partial}{\partial \theta} Q_n(\hat{\theta}_n)$$

$$= n^{-1/2}\frac{\partial}{\partial \boldsymbol{\theta}}Q_n(\boldsymbol{\theta}^0)$$
$$+ n^{-1}[V_n + T_n(\boldsymbol{\theta}^*)]n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0). \qquad (7.1.12)$$

其中

$$\frac{\partial}{\partial \boldsymbol{\theta}}Q_n(\boldsymbol{\theta}^0) = \frac{\partial}{\partial \boldsymbol{\theta}}Q_n(\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}.$$

由于当 $n\to\infty$ 时，

$$n^{-1}(V_n + T_n(\boldsymbol{\theta}^*)) \to 2V,\ \text{a.s.},$$

故由(7.1.12)式知，

$$n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0)\ \text{和}\ -(2V)^{-1}n^{-1/2}\frac{\partial}{\partial \boldsymbol{\theta}}Q_n(\boldsymbol{\theta}^0)$$

有相同的极限分布. 对任意非零向量 $\boldsymbol{c}=(c_1,\cdots,c_p)^\tau$，由(7.1.1)式，可以验证

$$n^{1/2}\boldsymbol{c}^\tau\frac{\partial}{\partial \boldsymbol{\theta}}Q_n(\boldsymbol{\theta}^0)$$
$$= -2n^{-1/2}\sum_{k=2}^{n}\left[\sum_{i=1}^{p}c_i\left\{\frac{\partial}{\partial \theta_i}E\theta\{x_k|\mathscr{F}_{1,k-1}^x\}\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\right]$$
$$\times\{x_k - E_{\boldsymbol{\theta}^0}\{x_k|\mathscr{F}_{1,k-1}^x\}\}. \qquad (7.1.13)$$

显然，

$$\sum_{k=1}^{p}c_i\left\{\frac{\partial}{\partial \theta_i}E_{\boldsymbol{\theta}}\{x_k|\mathscr{F}_{1,k-1}^x\}\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}$$

是 $\mathscr{F}_{1,k-1}^x$ 可测的，从而由(7.1.13)式易见

$$\left\{\boldsymbol{c}^\tau\frac{\partial}{\partial \boldsymbol{\theta}}Q_n(\boldsymbol{\theta}^0),\mathscr{F}_{1,n}^x,n\geqslant 1\right\}$$

是鞅. 如果当 $n\to\infty$ 时，有

$$\frac{1}{2}n^{-1/2}\boldsymbol{c}^\tau\frac{\partial}{\partial \boldsymbol{\theta}}Q_n(\boldsymbol{\theta}^0)\xrightarrow{d}N(0,\boldsymbol{c}^\tau W\boldsymbol{c}),$$

其中 W 是 $p\times p$ 协方差阵，于是，当 $n\to\infty$ 时，有

$$\frac{1}{2}n^{-1/2}\frac{\partial}{\partial \boldsymbol{\theta}}Q_n(\boldsymbol{\theta}^0)\xrightarrow{d}N(0,W).$$

从而有如下定理，其证明见文献[56]和[71]，或[106].

定理 7.1.2 假设定理 7.1.1 的条件成立. 进一步假定

$$\frac{1}{2}n^{-1/2}\frac{\partial}{\partial \boldsymbol{\theta}}Q_n(\boldsymbol{\theta}^0)\xrightarrow{d}N(0,W),\ \text{当}\ n\to\infty,$$

其中 W 是 $p\times p$ 正定对称矩阵，则

$$n^{1/2}(\hat{\theta}_n - \theta^0) \xrightarrow{d} N(0, V^{-1}WV^{-1}),\ 当\ n \to \infty.$$

下面讨论一些 NLAR 模型参数的 CLSE 和 LSE.

例 7.1.3　考虑如下 NLAR(1)模型

$$x_t = \varphi(x_{t-1}, \theta) + \varepsilon_t, \tag{7.1.14}$$

其中 $\varphi(\cdot, \theta)$关于 θ 有二阶连续导数. $\{x_1, \cdots, x_n\}$是来自 NLAR(1)模型(7.1.14)的样本,且$\{\varepsilon_1, \cdots, \varepsilon_n\}$与 x_0 相互独立. 由(7.1.14)式有

$$Q_n(\theta) = \sum_{j=2}^{n}[x_j - \varphi(x_{j-1}, \theta)]^2.$$

且 CLSE-$\hat{\theta}_n$ 满足方程

$$\sum_{j=2}^{n}[x_j - \varphi(x_{j-1}, \theta)]\frac{\mathrm{d}}{\mathrm{d}\theta}\varphi(x_{j-1}, \theta) = 0. \tag{7.1.15}$$

注意到

$$\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}Q_n(\theta)\right\}_{\theta=\theta^0} = -\sum_{j=2}^{n}\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\varphi(x_{j-1}, \theta)\right\}_{\theta=\theta^0}[x_j - \varphi(x_{j-1}, \theta_0)], \tag{7.1.16}$$

$$\begin{aligned}\left\{\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}Q_n(\theta)\right\}_{\theta=\theta^0} = &\sum_{j=2}^{n}\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\varphi(x_{j-1}, \theta)\right\}^2_{\theta=\theta^0} \\ &- \sum_{j=2}^{n}\left\{\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\varphi(x_{j-1}, \theta)\right\}_{\theta=\theta^0} \\ &\times [x_j - \varphi(x_{j-1}, \theta_0)].\end{aligned} \tag{7.1.17}$$

由(7.1.14)式、鞅的大数定律和鞅的中心极限定理,容易验证

$$\frac{1}{2}n^{-1/2}\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}Q_n(\theta)\right\}_{\theta=\theta^0} \xrightarrow{d} N(0, W),$$

且

$$V = W = E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\varphi(x_1, \theta)\right\}^2_{\theta=\theta^0}.$$

又易见,模型(7.1.14)满足定理 7.1.1 的条件. 所以根据定理 7.1.1 和定理 7.1.2,对方程(7.1.15)的解 $\hat{\theta}_n$,有

$$\hat{\theta}_n \to \theta^0,\ \text{a.s.},$$

$$n^{1/2}(\hat{\theta}_n - \theta^0) \xrightarrow{d} N(0, W^{-1}).$$

例如

$$\varphi(x, \theta) = \frac{\theta x}{1 + x^2},$$

即模型为

$$x_t = \frac{\theta x_{t-1}}{1 + x_{t-1}^2} + \varepsilon_t. \tag{7.1.18}$$

由(7.1.15)式可求得 θ 的 CLSE 为

$$\hat{\theta}_n = \left(\sum_{j=2}^{n} \frac{x_j x_{j-1}}{1 + x_{j-1}^2} \right) \Bigg/ \left(\sum_{j=2}^{n} \frac{x_{j-1}^2}{(1 + x_{j-1}^2)^2} \right) \tag{7.1.19}$$

且

$$W = E\left(\frac{x_1}{1 + x_1^2} \right)^2.$$

例 7.1.4 考虑如下的 AR(1)-MA(0)模型

$$x_t = (\theta + \eta_t) x_{t-1} + \varepsilon_t, \tag{7.1.20}$$

其中$\{\varepsilon_t\}$和$\{\eta_t\}$是相互独立的白噪声序列,并且 η_t 和 ε_t 都与 x_{t-1}独立. $E\varepsilon_t=0$, $E\varepsilon_t^2=\sigma_\varepsilon^2$, $E\varepsilon_t^4<\infty$, $E\eta_t=0$, $E\eta_t^2=\sigma_\eta^2$, $E\eta_t^3=0$, $E\eta_t^4=\mu_4<\infty$, $\theta^4+\sigma\theta^2\sigma_\eta^2+\mu_4<1$. 则由 § 4.4 的讨论知,模型(7.1.20)不仅是平稳的而且是宽平稳的,并且具有有穷的四阶矩.

给定样本 $x_1,\cdots,x_n$,由(7.1.1)式有

$$\begin{aligned} Q_n(\theta) &= \sum_{j=2}^{n} \{x_j - E_\theta\{x_j | \mathscr{F}_{1,j-1}^x\}\}^2 \\ &= \sum_{j=2}^{n} (x_j - \theta x_{j-1})^2. \end{aligned}$$

因此,θ 的 CLSE 为

$$\hat{\theta}_n = \left(\sum_{j=2}^{n} x_{j-1}^2 \right)^{-1} \left(\sum_{j=2}^{n} x_j x_{j-1} \right). \tag{7.1.21}$$

令

$$u_t = \eta_t x_{t-1} + \varepsilon_t,$$

则

$$E\{u_t^2 | \mathscr{F}_{1,t-1}^x\} = x_{t-1}^2 \sigma_\eta^2 + \sigma_\varepsilon^2.$$

再令

$$\hat{u}_t = x_t - \hat{\theta}_n x_{t-1}, \quad \xi_t = \hat{u}_t^2 - x_{t-1}^2 \sigma_\eta^2 - \sigma_\varepsilon^2,$$

那么关于 σ_η^2 和 σ_ε^2 极小化

$$\overline{Q}_n(\sigma_\eta^2, \sigma_\varepsilon^2) = \sum_{j=2}^{n} \xi_j^2,$$

可得到 σ_η^2 和 σ_ε^2 的 CLSE 分别为

$$\hat{\sigma}_\eta^2 = \left\{ \sum_{j=2}^{n} (x_{j-1}^2 - \overline{x}^2)^2 \right\}^{-1}$$

$$\times \left\{ \sum_{j=2}^{n} (\hat{u}_j^2 - \overline{u}^2)(x_{j-1}^2 - \overline{x}^2) \right\} \qquad (7.1.22)$$

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n-1} \sum_{j=2}^{n} \hat{u}_j^2 - \hat{\sigma}_\eta^2 \overline{x}^2, \qquad (7.1.23)$$

其中

$$\overline{x}^2 = \frac{1}{n} \sum_{j=1}^{n} x_j^2,$$

$$\overline{u}^2 = \frac{1}{n-1} \sum_{t=2}^{n} \hat{u}_t^2.$$

在模型(7.1.20)的假定下,可以证明 $\hat{\theta}_n$、$\hat{\sigma}_\varepsilon^2$ 和 $\hat{\sigma}_\eta^2$ 都是真值 θ^0、σ_ε^2 和 σ_η^2 的强相容估计,且具有渐近正态性. 详细证明见文献[82].

例 7.1.5 考虑如下门限自回归模型

$$x_t = \sum_{j=1}^{l} \left\{ a_{j0} + \sum_{i=1}^{p} a_{ji} x_{t-i} + \varepsilon_{jt} \right\} I(x_{t-d} \in R_j), \qquad (7.1.24)$$

其中自回归系数

$$\boldsymbol{\theta} = (\boldsymbol{a}_1^\tau, \cdots, \boldsymbol{a}_l^\tau)^\tau,$$

$$\boldsymbol{a}_j = (a_{j0}, a_{j1}, \cdots, a_{jp})^\tau \quad (j = 1, \cdots, l),$$

$\{\varepsilon_{jt}\}$ $(j=1,\cdots,l)$满足例 4.3.15 的条件, $E\varepsilon_{jt}=0$, $E\varepsilon_{jt}^2=\sigma_j^2$, $E\varepsilon_{jt}^4<\infty$. 当 $i\neq j$ 时, $\boldsymbol{a}_i\neq\boldsymbol{a}_j$, $R_1=(-\infty, r_1]$, $R_j=(r_{j-1}, r_j]$ $(j=2,\cdots,l-1)$, $R_l=(r_{l-1},\infty)$. 假定参数 p、d、r_j 及 l 均已知,且 $0<d\leqslant p$.

如果由(7.1.24)式获得观测样本$\{x_1,\cdots,x_n\}$,则可通过 CLSE 来估计真参数 $\boldsymbol{\theta}^0=(\boldsymbol{a}_{i0}^\tau,\cdots,\boldsymbol{a}_{i0}^\tau)^\tau$,其中 $\boldsymbol{a}_{i0}^\tau=(a_{i0}^0, a_{i1}^0,\cdots,a_{ip}^0)^\tau$ $(i=1,\cdots,l)$及 σ_{0j}^2 $(j=1,\cdots,l)$,即关于 $\boldsymbol{\theta}$ 极小化如下的残差平方和:

$$Q_n(\boldsymbol{\theta}) = \sum_{t=p+1}^{n} \{x_t - E_{\boldsymbol{\theta}}\{x_t | \mathscr{F}_{1,t-1}^{x}\}\}^2$$

$$= \sum_{t=p+1}^{n} \left\{ x_t - \sum_{j=1}^{l} \left[a_{j0} + \sum_{i=1}^{p} a_{ji} x_{t-i} \right] \right.$$

$$\left. \times I(x_{t-d} \in R_j) \right\}^2. \qquad (7.1.25)$$

记 $\boldsymbol{\theta}^0$ 的 CLSE 为 $\hat{\boldsymbol{\theta}}_n$,则 $\hat{\boldsymbol{\theta}}_n$ 满足如下方程:

$$(X_n^\tau X_n)\hat{\theta}_n = X_n^\tau \boldsymbol{y}_n,$$

其中

$$y_n = (x_{p+1}, \cdots, x_n)^\tau,$$

$$x_{ij}^\tau = (1, x_{i-1}, \cdots, x_{i-p}) I(x_{t-d} \in R_j),$$

$$X_n = \begin{pmatrix} x_{p+1\,1}^\tau & x_{p+1\,2}^\tau & \cdots & x_{p+1\,l}^\tau \\ x_{p+2\,1}^\tau & x_{p+2\,2}^\tau & \cdots & x_{p+2\,l}^\tau \\ \cdots & \cdots & \cdots & \cdots \\ x_{n\,1}^\tau & x_{n\,2}^\tau & \cdots & x_{n\,l}^\tau \end{pmatrix}.$$

由文献[127]中的引理 4 和定理 4 知,存在正定对称矩阵 H,使得

$$\frac{1}{n} X_n^\tau X_n \to H > 0, \text{a.s.}, \quad 当\ n \to \infty\ 时.$$

于是当 n 适当大以后,$X_n^\tau X_n$ 是可逆的,从而有

$$\hat{\theta}_n = (X_n^\tau X_n)^{-1} X_n^\tau y_n. \tag{7.1.26}$$

令 $x_{(1)} \leqslant x_{(2)} \leqslant \cdots \leqslant x_{(n)}$ 是样本 $\{x_1, \cdots, x_n\}$ 的次序统计量,则 $Q_n(\hat{\theta}_n)$ 可分解成 l 个残差平方和之和,即

$$Q_n(\hat{\theta}_n) = \sum_{j=1}^{l} Q_{jn}(\hat{\theta}_n), \tag{7.1.27}$$

其中

$$Q_{jn}(\hat{\theta}_n) = \sum_{x_{(i)-d} \in R_j} \{x_{(i)} - E_{\hat{\theta}_n}\{x(i) \mid \mathscr{F}_{1(i)-1}^x\}\}^2,$$

式中 $\mathscr{F}_{1(i)-1}^x = \sigma(x_s, 1 \leqslant s \leqslant (i)-1)$. 以

$$\hat{\sigma}_{jn}^2 = Q_{jn}(\hat{\theta}_n)/n_j \quad (j = 1, \cdots, l) \tag{7.1.28}$$

作为 σ_{0j}^2 的 CLSE,其中

$$n_j = \sum_{j=p+1}^{n} I(x_{t-d} \in R_j) \quad (j = 1, \cdots, l).$$

在模型(7.1.24)的假定下,估计量 $\hat{\theta}_n$ 和 $\hat{\sigma}_{jn}^2$ 是真值 θ^0 和 σ_{0j}^2 的强相容估计,且具有渐近正态性,详细证明见文献[127].

在下一例中,我们讨论门限自回归模型(7.1.24)中门限 r_i 和延迟量 d 的估计. 为简便计,仅讨论 $l=2$ 的情形.

例 7.1.6 考虑如下门限自回归模型:

$$x_t = \begin{cases} a_{10} + a_{11}x_{t-1} + \cdots + a_{1p}x_{t-p} + \varepsilon_{1t}, \\ \qquad 当\ x_{t-d} \leqslant r; \\ a_{20} + a_{21}x_{t-1} + \cdots + a_{2p}x_{t-p} + \varepsilon_{2t}, \\ \qquad 当\ x_{t-d} > r, \end{cases} \tag{7.1.29}$$

其中自回归系数和$\{\varepsilon_{it}\}$ $(i=1,2)$满足例 7.1.5 的条件，p 是已知的，且 $d\leqslant p$. 记 $\boldsymbol{\theta}^0=(\boldsymbol{a}_{10}^{\tau},\boldsymbol{a}_{20}^{\tau},r_0,d_0)^{\tau}$. 给定样本$\{x_1,\cdots,x_n\}$，按如下四步求出 $\boldsymbol{\theta}^0$ 和 σ_{0j}^2的 CLSE 估计 $\hat{\boldsymbol{\theta}}^0$ 和 $\hat{\sigma}_{0j}^2$ $(j=1,2)$：

第一步：对固定的 r 和 d，关于 $\boldsymbol{a}_1$ 和 $\boldsymbol{a}_2$ 极小化(7.1.25)式，得到 $\boldsymbol{a}_{10}$和 $\boldsymbol{a}_{20}$的 CLSE：$\hat{\boldsymbol{a}}_{1n}(r,d)$和 $\hat{\boldsymbol{a}}_{2n}(r,d)$，其表示式由(7.1.26)式给出.

第二步：对固定的 d，关于 r 极小化

$$
\begin{aligned}
Q_n(\hat{\boldsymbol{\theta}}_{1n})&=\sum_{t=p+1}^{n}\{x_t-E_{\hat{\theta}_{1n}}\{x_t\,|\,\mathscr{F}_{1t-1}^{x}\}\}^2\\
&=\sum_{t=p+1}^{n}\{x_t-[\hat{\boldsymbol{a}}_{1n}^{\tau}(r,d)\ (1,x_{t-1},\cdots,x_{t-p})^{\tau}I(x_{t-d}\leqslant r)\\
&\qquad+\hat{\boldsymbol{a}}_{2n}^{\tau}(r,d)\ (1,x_{t-1},\cdots,x_{t-p})^{\tau}I(x_{t-d}>r)]\}^2,
\end{aligned}
$$

其中 $\hat{\boldsymbol{\theta}}_{1n}=(\hat{\boldsymbol{a}}_{1n}^{\tau}(r,d)\ ,\hat{\boldsymbol{a}}_{2n}^{\tau}(r,d)\ ,r,d)^{\tau}$. 记其极小解为 $\hat{r}_n(d)$.

第三步：关于 d 极小化

$$
\begin{aligned}
Q_n(\hat{\boldsymbol{\theta}}_{2n})&=\sum_{t=p+1}^{n}\{x_t-E_{\hat{\theta}_{2n}}\{x_t\,|\,\mathscr{F}_{1t-1}^{x}\}\}^2\\
&=\sum_{t=p+1}^{n}\{x_t-[\hat{\boldsymbol{a}}_{1n}^{\tau}(\hat{r}_n(d),d)(1,x_{t-1},\cdots,x_{t-p})^{\tau}\\
&\qquad\times I(x_{t-d}\leqslant\hat{r}_n(d))\\
&\qquad+\hat{\boldsymbol{a}}_{2n}^{\tau}(\hat{r}_n(d),d)(1,x_{t-1},\cdots,x_{t-p})^{\tau}I(x_{t-d}>\hat{r}_n(d))]\}^2,
\end{aligned}
$$

记其极小解为 $\hat{d}_n$，并以 $\hat{r}_n(\hat{d}_n)$、$\hat{\boldsymbol{a}}_{1n}(\hat{r}_n(\hat{d}_n),\hat{\boldsymbol{d}}_n)$、$\hat{\boldsymbol{a}}_{2n}(\hat{r}_n(\hat{d}_n),\hat{d}_n)$分别作为 r_0、$\boldsymbol{a}_{10}$、$\boldsymbol{a}_{20}$的 CLSE.

第四步：按(7.1.28)式的方法，把 $Q_n(\hat{\theta}_n)$分解成两个残差平方和之和，即

$$
Q_n(\hat{\theta}_n)=Q_{1n}(\hat{\theta}_n)+Q_{2n}(\hat{\theta}_n),
$$

其中

$$
\begin{aligned}
\hat{\boldsymbol{\theta}}_n&=(\hat{\boldsymbol{a}}_{1n}^{\tau}(\hat{r}_n(\hat{d}_n),\hat{d}_n),\hat{\boldsymbol{a}}_{2n}^{\tau}(\hat{r}_n(\hat{d}_n),\hat{d}_n),\hat{r}_n(\hat{d}_n),\hat{d}_n)^{\tau},\\
Q_{1n}(\hat{\boldsymbol{\theta}}_n)&=\sum_{x_{(i)-\hat{d}_n}\leqslant\hat{r}_n(\hat{d}_n)}\{x_{(i)}-E_{\hat{\theta}_n}\{x_{(i)}\,|\,\mathscr{F}_{1(i)-1}^{x}\}\}^2,\\
Q_{2n}(\hat{\boldsymbol{\theta}}_n)&=Q_n(\hat{\boldsymbol{\theta}}_n)-Q_{1n}(\hat{\boldsymbol{\theta}}_n).
\end{aligned}
$$

以如下的估计量

$$
\hat{\sigma}_{in}^2=Q_{in}(\hat{\boldsymbol{\theta}}_n)/n_i\quad(i=1,2)
$$

作为 σ_i^2 的 CLSE，其中

$$n_1 = \sum_{t=p+1}^{n} I(x_{t-\hat{d}_n} \leqslant \hat{r}_n(\hat{d}_n)), \quad n_2 = n - p - n_1.$$

在模型(7.1.29)的假定下,Chan(1992)证明了 $\hat{\theta}_n$ 是 $\hat{\theta}^0$ 的强相容估计,$\hat{\sigma}_{in}^2$ 是 σ_{0i}^2 的强相容估计,且 $n^{1/2}(\hat{\boldsymbol{a}}_n - \boldsymbol{a}_0)$ 具有渐近正态性,其中

$$\hat{\boldsymbol{a}}_n = (\hat{\boldsymbol{a}}_{1n}^{\tau}(\hat{r}_n(\hat{d}_n), \hat{d}_n),\ \hat{\boldsymbol{a}}_{2n}^{\tau}(\hat{r}_n(\hat{d}_n), \hat{d}_n))^{\tau},\ \boldsymbol{a}_0 = (\boldsymbol{a}_{10}^{\tau}, \boldsymbol{a}_{20}^{\tau})^{\tau}.$$

在实际数据分析中,固定 d,关于 r 求 $Q_n(\hat{\theta}_{1n})$ 的极小解,可先根据数据直观分析,预先确定一些门限候选值,在这些值处比较 $Q_n(\hat{\theta}_m)$ 的大小,取使 $Q_n(\hat{\theta}_{1n})$ 最小的候选值作为 $\hat{r}_n(d)$. 例如先将数据从小到大重新排列,以位于数据量 25%,35%,45%,55%,65%和 75%处的 x_t 的取值作为门限候选值.

最后来讨论双线性模型参数的 CLSE.

例 7.1.7 考虑如下双线性模型

$$x_t = \sum_{j=1}^{p} a_j x_{t-j} + b\varepsilon_{t-1}x_{t-2} + \varepsilon_t, \tag{7.1.30}$$

其中$\{\varepsilon_t\}$为正态白噪声序列,$\varepsilon_t \sim N(0,\sigma^2)$,$\varepsilon_t$ 与$\{x_s, s<t\}$独立. 类似于 § 4.4 和 § 4.5 的讨论,记

$$A = \begin{pmatrix} a_1 & \cdots & a_{p-1} & a_p \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix},$$

$$B = \begin{pmatrix} 0 & b & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

如果

$$\rho(A^{\otimes 2} + \sigma^2 B^{\otimes 2}) < 1, \tag{7.1.31}$$

则$\{x_t\}$是平稳的,且 $E|x_t|^2 < \infty$. 如果

$$\rho(\Lambda) < 1, \tag{7.1.32}$$

其中矩阵 Λ 的定义见 § 4.5,则 $E|x_t|^4 < \infty$. 令

$$r_s = E(x_t - \mu)(x_{t+s} - \mu),\ \mu = Ex_t,$$

$$v_t = x_t - a_1 x_{t-1} - \cdots - a_p x_{t-p},$$

及

$$\mathscr{F}_t^{\varepsilon} = \sigma(\varepsilon_s, s \leqslant t),$$

则不难验证

$$r_s = \sum_{j=1}^{p} a_j r_{s-j},\ s \geqslant 2. \tag{7.1.33}$$

$$E\{v_t v_{t-s} | \mathscr{F}^{\varepsilon}_{t-2}\} = \begin{cases} \sigma^2 + b^2\sigma^2 x_{t-2}^2, & 当\ s = 0; \\ b\sigma^2 x_{t-2}, & 当\ s = 1; \\ 0, & 当\ s \geqslant 2. \end{cases} \tag{7.1.34}$$

于是可按如下步骤来求参数 $\boldsymbol{\theta}=(a_1,\cdots,a_p,b,\sigma^2)^{\tau}$ 的估计：

第一步：记 $\hat{\boldsymbol{\gamma}}(p)=(\hat{\gamma}_2,\cdots,\hat{\gamma}_{p+1})^{\tau}$，$\hat{\boldsymbol{a}}_n=(\hat{a}_{1n},\cdots,\hat{a}_{pn})^{\tau}$，

$$\hat{\boldsymbol{\Gamma}}_p = \begin{pmatrix} \hat{\gamma}_1 & \hat{\gamma}_0 & \cdots & \hat{\gamma}_{-p+2} \\ \hat{\gamma}_2 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{-p+3} \\ \cdots & \cdots & \cdots & \cdots \\ \hat{\gamma}_p & \hat{\gamma}_{p-1} & \cdots & \hat{\gamma}_1 \end{pmatrix},$$

其中

$$\hat{\gamma}_s = \frac{1}{n}\sum_{t=1}^{n-s}(x_t - \hat{\mu}_n)(x_{t+s} - \hat{\mu}_n) \quad (s = 0,1,\cdots,n-1),$$

$$\hat{\mu}_n = \frac{1}{n}\sum_{t=1}^{n} x_t.$$

由(7.1.33)式得 Yule-Walker 方程

$$\hat{\boldsymbol{\Gamma}}_p \hat{\boldsymbol{a}}_n = \hat{\boldsymbol{\gamma}}(p).$$

仿照常学将等的文献[132]中引理 2.1，可证 $\hat{\boldsymbol{\Gamma}}_p$ 是 a.s. 可逆的. 从而解上述矩阵方程得

$$\hat{\boldsymbol{a}}_n = \hat{\boldsymbol{\Gamma}}_p^{-1}\hat{\boldsymbol{\gamma}}(p). \tag{7.1.35}$$

第二步：令

$$\hat{v}_t = x_t - \hat{a}_{1n}x_{t-1} - \cdots - \hat{a}_{pn}x_{t-p} \quad (t = p+1,\cdots,n).$$

记 $\beta_1=\sigma^2$，$\beta_2=b^2\sigma^2$. 利用(7.1.34)式，关于 β_1、β_2，极小化

$$\begin{aligned} Q_{1n}(\beta_1,\beta_2) &= \sum_{t=p+1}^{n}\{v_t^2 - E_{\beta_1,\beta_2}\{v_t^2 | \mathscr{F}^{\varepsilon}_{t-2}\}\}^2 \\ &= \sum_{t=p+1}^{n}[\hat{v}_t^2 - (\beta_1 + \beta_2 x_{t-2}^2)]^2. \end{aligned}$$

记其极小解为 $\hat{\beta}_{1n}$和 $\hat{\beta}_{2n}$，通过解如下方程组：

$$\begin{cases} \sum\limits_{t=p+1}^{n}[\hat{v}_t^2 - (\beta_1 + \beta_2 x_{t-2}^2)] = 0; \\ \sum\limits_{t=p+1}^{n}[\hat{v}_t^2 - (\beta_1 + \beta_2 x_{t-2}^2)]x_{t-2}^2 = 0, \end{cases}$$

可得

$$\hat{\beta}_{2n} = \frac{\sum_{t=p+1}^{n} (\hat{v}_t^2 - \bar{v}^2)(x_{t-2}^2 - \bar{v}^2)}{\sum_{t=p+1}^{n} (x_{t-2}^2 - \bar{x}^2)^2},$$

其中

$$\bar{x}^2 = \frac{1}{n-p} \sum_{t=p+1}^{n} x_{t-2}^2, \quad \bar{v}^2 = \frac{1}{n-p} \sum_{t=p+1}^{n} \hat{v}_t^2,$$

$$\hat{\beta}_{1n} = \bar{v}^2 - \hat{\beta}_{2n} \bar{x}^2.$$

第三步：再记 $\beta_3 = b\sigma^2$，关于 β_3 极小化

$$\begin{aligned} Q_{2n}(\beta_3) &= \sum_{t=p+1}^{n} \{\hat{v}_t \hat{v}_{t-1} - E_{\beta_3}\{v_t v_{t-1} | \mathscr{F}_{t-2}^{t}\}\}^2 \\ &= \sum_{t=p+1}^{n} (\hat{v}_t \hat{v}_{t-1} - \beta_3 x_{t-2})^2. \end{aligned}$$

记其极小解为 $\hat{\beta}_{3n}$，则

$$\hat{\beta}_{3n} = \frac{\sum_{t=p+1}^{n} \hat{v}_t \hat{v}_{t-1} x_{t-2}}{\sum_{t=p+1}^{n} x_{t-2}^2}.$$

这样，由(7.1.34)式知，b 和 σ^2 的估计分别为

$$\hat{b}_n = \hat{\beta}_{3n} / \hat{\beta}_{1n}, \quad \hat{\sigma}_n^2 = \hat{\beta}_{1n}. \tag{7.1.36}$$

对于更一般的双线性模型

$$x_t = \sum_{j=1}^{p} a_j x_{t-j} + \sum_{j=1}^{q} c_j \varepsilon_{t-j} + \sum_{i=1}^{Q} \sum_{j=r}^{P} b_{ij} \varepsilon_{t-i} x_{t-j} + \varepsilon_t, \tag{7.1.37}$$

其中 $r = \max\{q, Q\} + 1$，$\{\varepsilon_t\}$ 为正态白噪声序列，$\varepsilon_t \sim N(0, \sigma^2)$. 参数

$$\boldsymbol{\theta} = (a_1, \cdots, a_p, c_1, \cdots, c_q, b_{ij}, i = 1, \cdots, Q;\ j = r, \cdots, P)^{\tau}$$

满足§4.4和§4.5中关于双线性模型有平稳解和高阶矩的要求. 对于模型(7.1.37)，$\boldsymbol{\theta}$ 的 CLSE 比模型(7.1.30)更复杂，具体算法见文献[53].

利用文献[53]中的结果，可以证明在适当的条件下，由(7.1.35)和(7.1.36)给出的模型(7.1.30)的参数 $\boldsymbol{\theta}$ 的 CLSE 是强相容的，且具有渐近正态性. 对其他形式的双线性模型的 CLSE 及其渐近性质，还有待研究.

二、矩估计

在线性时间序列的参数估计中，特别是 LARMA 模型的参数估计中，常用矩估计法求模型中未知参数的估计. 这是因为线性时间序列模型

的参数可由该序列的二阶矩唯一确定. 例如对 LAR(p)模型

$$x_t = \varphi_1 x_{t-1} + \cdots + \varphi_p x_{t-p} + \varepsilon_t, \tag{7.1.38}$$

其中$\{\varepsilon_t\}$为白噪声序列,满足(3. 2. 2)式,具有有穷四阶矩,且参数$\boldsymbol{\varphi}=(\varphi_1,\cdots,\varphi_p)^\tau$满足(1. 2. 5)式,记$\gamma_k = Ex_t x_{t+k}$,有如下熟知的Yule-Walker方程

$$\gamma_k = \varphi_1 \gamma_{k-1} + \cdots + \varphi_p \gamma_{k-p}, k \geqslant 1. \tag{7.1.39}$$

对$k=1,2,\cdots,p$,在上式中用样本协方差$\hat{\gamma}_k$代替γ_k,并解方程

$$\begin{pmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{p-1} \\ \hat{\gamma}_1 & \hat{\gamma}_0 & \cdots & \hat{\gamma}_{p-2} \\ \cdots & \cdots & \cdots & \cdots \\ \hat{\gamma}_{p-1} & \hat{\gamma}_{p-2} & \cdots & \hat{\gamma}_0 \end{pmatrix} \begin{pmatrix} \hat{\varphi}_1 \\ \hat{\varphi}_2 \\ \vdots \\ \hat{\varphi}_p \end{pmatrix} = \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \vdots \\ \hat{\gamma}_p \end{pmatrix},$$

可得参数$\boldsymbol{\varphi}$的 Yule-Walker 估计.

对非线性时间序列,只有极少数的模型可用矩估计法求其参数估计,例如双线性模型、双重线性模型、拟线性模型等. 而且对有些非线性模型,仅用二阶矩还不能唯一地确定模型的参数,常需借助于三阶矩或四阶矩,这在下文将有所讨论. 现在先来讨论一个简单的双线性模型.

考虑双线性模型

$$x_t = b\varepsilon_{t-1}x_{t-1} + \varepsilon_t, \tag{7.1.40}$$

其中$\{\varepsilon_t\}$是正态白噪声序列,$\varepsilon_t \sim N(0,\sigma^2)$,且$|b\sigma|<1$. 则(7. 1. 40)式有平稳解. 记$\lambda=b\sigma$,不难验证

$$\mu = Ex_t = b\sigma^2, \tag{7.1.41}$$

$$\gamma_k = \begin{cases} \sigma^2(1+\lambda^2+\lambda^4)/(1-\lambda^2), & \text{当 } k=0; \\ \sigma^2\lambda^2, & \text{当 } |k|=1; \\ 0, & \text{当 } |k|\geqslant 2, \end{cases} \tag{7.1.42}$$

其中$\gamma_k = E(x_t-\mu)(x_{t+k}-\mu)$. 事实上,由(7. 1. 40)式有

$$Ex_t^2 = E(\varepsilon_t + b\varepsilon_{t-1}x_{t-1})^2 = \sigma^2 + b^2 E\varepsilon_{t-1}^2 x_{t-1}^2. \tag{7.1.43}$$

而

$$\begin{aligned} E\varepsilon_t^2 x_t^2 &= E\varepsilon_t^2(\varepsilon_t + b\varepsilon_{t-1}x_{t-1})^2 \\ &= E\varepsilon_t^4 + b^2 E\varepsilon_t^2\varepsilon_{t-1}^2 x_{t-1}^2 \\ &= 3\sigma^4 + b^2\sigma^2 E\varepsilon_{t-1}^2 x_{t-1}^2, \end{aligned}$$

由上式,并注意$\{x_t\}$的平稳性,可得

$$E\varepsilon_t^2 x_t^2 = \frac{3\sigma^4}{1-\lambda^2}.$$

从而再由(7.1.43)式,有

$$Ex_t^2 = \frac{\sigma^2(1-\lambda^2)+3\sigma^2\lambda^2}{1-\lambda^2}. \tag{7.1.44}$$

又

$$\begin{aligned} Ex_t &= E(\varepsilon_t + b\varepsilon_{t-1}x_{t-1}) = b\varepsilon_{t-1}x_{t-1} \\ &= bE\varepsilon_{t-1}(\varepsilon_{t-1} + b\varepsilon_{t-2}x_{t-2}) \\ &= bE\varepsilon_{t-1}^2 = b\sigma^2. \end{aligned}$$

这样,根据上式及(7.1.44)式,可得

$$\gamma_0 = Ex_t^2 - (Ex_t)^2 = \frac{\sigma^2(1+\lambda^2+\lambda^4)}{1-\lambda^2}.$$

用类似的方法可证(7.1.42)式中的其他两式.

把 $\lambda^2=\gamma_1/\sigma^2$ 代入(7.1.42)式的第一个等式中,可解得

$$\sigma^2 = \frac{1}{2}[(\gamma_0-\gamma_1) \pm (\gamma_0^2 - 6\gamma_1\gamma_0 - 3\gamma_1^2)^{1/2}]. \tag{7.1.45}$$

利用(7.1.42)式,不难验证

$$\begin{aligned} \gamma_0 - \gamma_1 &= \frac{\sigma^2 + 2\sigma^2\lambda^4}{1-\lambda^2}, \\ (\gamma_0^2 - 6\gamma_0\gamma_1 - 3\gamma_1^2)^{1/2} &= \frac{\sigma^2|2\lambda^4+2\lambda^2-1|}{1-\lambda^2}. \end{aligned} \tag{7.1.46}$$

由以上两式可知

$$\gamma_0 - \gamma_1 > (\gamma_0^2 - 6\gamma_0\gamma_1 - 3\gamma_1^2)^{1/2}.$$

从而只要(7.1.45)式有实数解,就总有 $\sigma^2>0$. 为使

$$\sigma^2 = \frac{1}{2}[(\gamma_0-\gamma_1) + (\gamma_0^2 - 6\gamma_1\gamma_0 - 3\gamma_1^2)^{1/2}] \tag{7.1.47}$$

有实数解,只需

$$\gamma_0^2 - 6\gamma_1\gamma_0 - 3\gamma_1^2 > 0.$$

而根据(7.1.46)式,为使上式成立,只需

$$2\lambda^4 + 2\lambda^2 - 1 < 0.$$

由上式解得

$$|\lambda| < \{(\sqrt{3}-1)/2\}^{1/2} = \lambda_0 \approx 0.605. \tag{7.1.48}$$

对模型(7.1.40),假定(7.1.47)和(7.1.48)两式满足.令

$$\hat{\mu}_n = \frac{1}{n}\sum_{t=1}^{n} x_t,$$

$$\hat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k}(x_t - \hat{\mu}_n)(x_{t+k} - \hat{\mu}_n) \quad (k = 0,1),$$

则由(7.1.42)和(7.1.47)两式得到 b 和 σ^2 的矩估计：

$$\begin{cases} \hat{\sigma}_n^2 = \dfrac{1}{2}[(\hat{\gamma}_0 - \hat{\gamma}_1) + (\hat{\gamma}_0^2 - 6\hat{\gamma}_0\hat{\gamma}_1 - 3\hat{\gamma}_1^2)^{1/2}]; \\ \hat{b}_n = \dfrac{\hat{\mu}_n}{\hat{\sigma}_n^2}. \end{cases} \tag{7.1.49}$$

在适当的条件下，Kim 等(见文献[70])证明了：由(7.1.49)式给出的模型(7.1.40)参数 (b,σ^2) 的矩估计是弱相容的，且具有渐近正态性.

下面考虑比(7.1.40)更一般的双线性模型

$$x_t + \sum_{j=1}^{p} a_j x_{t-j} = \varepsilon_t + \sum_{j=1}^{p} b_j x_{t-j}\varepsilon_{t-1}, \tag{7.1.50}$$

其中 $\{\varepsilon_t\}$ 仍为正态白噪声序列，$\varepsilon_t \sim N(0,\sigma^2)$. 且

$$\rho(A^{\otimes 2} + \sigma^2 B^{\otimes 2}) < 1,$$

其中 A 如模型(7.1.30)中所定义；

$$B = \begin{pmatrix} b_1 & \cdots & b_p \\ 0 & \cdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \cdots & 0 \end{pmatrix},$$

则类似于(7.1.42)的证明，可得到

$$\begin{cases} \gamma_1 = -\displaystyle\sum_{j=1}^{p} a_j \gamma_{j-1} + \sum_{j=1}^{p} b_j \sigma^2 \mu, \\ \gamma_s = -\displaystyle\sum_{j=1}^{p} a_j \gamma_{s-j}, \quad s \geqslant 2. \end{cases} \tag{7.1.51}$$

由此可见，当 $p>2$ 时，关于 $b_1,\cdots,b_p$ 的信息只包含在(7.1.51)的第一个方程中，并以 $b_1+\cdots+b_p$ 的形式出现，从而用二阶矩的差分方程(7.1.51)无法得到参数 $\theta=(a_1,\cdots,a_p,b_1,\cdots,b_p)^\tau$ 和 σ^2 的全部估计.

记 $\gamma_{s_1,s_2} = E(x_t-\mu)(x_{t+s_1}-\mu)(x_{t+s_2}-\mu)$，则可以证明

$$\begin{cases} \gamma_{s,s} = \displaystyle\sum_{j=1}^{p}\sum_{k=1}^{p} \alpha(j,k)\gamma_{s-j,s-k} \\ \qquad + 2\displaystyle\sum_{j=1}^{p} b_j\sigma^2\left(-a_1 + \sum_{k=1}^{p} b_k\mu\right)\gamma_{s-j}, \quad s \geqslant 2; \\ \gamma_{s,s+1} = -\displaystyle\sum_{j=1}^{p} a_j \gamma_{s,s-j+1} + \sum_{j=1}^{p} b_j \sigma^2 \gamma_{s-j+1}, \quad s \geqslant 1. \end{cases} \tag{7.1.52}$$

其中 $\alpha(j,k)=a_j a_k+b_j b_k\sigma^2$.(7.1.52)式的证明方法与(7.1.42)式类似，但其计算繁琐冗长，故略去，读者可参阅文献[94].

如果 σ^2 已知，则在(7.1.51)式的第二个方程中取 $s=2,3,\cdots,p+1$；在(7.1.52)式的第二个方程中取 $s=1,2,\cdots,p$，得到如下联立方程组：

$$\begin{cases}\gamma_2=-\sum_{j=1}^{p}a_j\gamma_{2-j},\\ \cdots\cdots\cdots\cdots\\ \gamma_{p+1}=-\sum_{j=1}^{p}a_j\gamma_{p+1-j},\\ \gamma_{1,2}=-\sum_{j=1}^{p}a_j\gamma_{1,2-j}+\sum_{j=1}^{p}b_j\sigma^2\gamma_{2-j},\\ \cdots\cdots\cdots\cdots\\ \gamma_{p,p+1}=-\sum_{j=1}^{p}a_j\gamma_{p,p+1-j}+\sum_{j=1}^{p}b_j\sigma^2\gamma_{p+1-j}.\end{cases}\tag{7.1.53}$$

在(7.1.53)式中用

$$\hat{\gamma}_k=\frac{1}{n}\sum_{t=1}^{n-k}(x_t-\hat{\mu}_n)(x_{t+k}-\hat{\mu}_n)$$
$$(k=0,1,\cdots,n-1),$$
$$\hat{\gamma}_{s_1,s_2}=\frac{1}{n}\sum_{t=1}^{n-\max(s_1,s_2)}(x_t-\hat{\mu}_n)(x_{t+s_1}-\hat{\mu}_n)(x_{t+s_2}-\hat{\mu}_n)$$
$$(s_1、s_2=0,1,\cdots,n-1)$$

代替 γ_k 和 γ_{s_1,s_2}，解此方程组可得参数 θ 的矩估计.

当 σ^2 未知时，可先由(7.1.53)式利用样本 $x_1,\cdots,x_n$ 来解出参数 $a_1,\cdots,a_p$，$c_1=b_1\sigma^2,\cdots,c_p=b_p\sigma^2$ 的矩估计 $\hat{a}_1,\cdots,\hat{a}_p,\hat{c}_1,\cdots,\hat{c}_p$，然后在(7.1.52)式的第一个方程中，取 $s=2,3,\cdots,p+2$，并用$\hat{\gamma}_s$、$\hat{\gamma}_{s_1,s_2}$代替相应的 γ_s、γ_{s_1,s_2}，可得到 $\alpha(j,k)$和 $d=-a_1+(b_1+\cdots+b_p)\mu$ 的矩估计 $\hat{\alpha}(j,k)$，$\hat{d}$. 再由

$$\hat{\alpha}(1,k)=\hat{a}_1\hat{a}_k+\hat{b}_k\hat{c}_1\quad(j=1,\cdots,p),$$
$$\hat{\sigma}^2=\hat{c}_1/\hat{b}_1$$

求得 $\hat{b}_k(k=1,\cdots,p)$和 $\hat{\sigma}^2$.

例 7.1.8 考虑如下双线性模型

$$x_t+ax_{t-1}=\varepsilon_t+bx_{t-2}\varepsilon_{t-1}.\tag{7.1.54}$$

其中$\{\varepsilon_t\}$是正态白噪声序列，满足(3.2.2)式，且具有有穷的四阶矩，而且

$a^2+\sigma^2b^2<1$. 由(7.1.53)式

$$\begin{cases}\hat{\gamma}_2=-\hat{a}\hat{\gamma}_1;\\ \hat{\gamma}_{1,2}=-\hat{a}\hat{\gamma}_{1,1}+\hat{c}\hat{\gamma}_1,\end{cases}$$

其中 $\hat{c}=\hat{b}\hat{\sigma}^2$,求解得到

$$\begin{cases}\hat{a}=-\hat{\gamma}_2/\hat{\gamma}_1;\\ \hat{c}=\dfrac{\hat{\gamma}_{1,2}+\hat{a}\hat{\gamma}_{1,1}}{\hat{\gamma}_1}.\end{cases}\tag{7.1.55}$$

又由

$$\begin{cases}\hat{\gamma}_{2,2}=\hat{\alpha}(1,1)\hat{\gamma}_{1,1}+2\hat{d}\hat{c}\hat{\gamma}_1;\\ \hat{\gamma}_{3,3}=\hat{\alpha}(1,1)\hat{\gamma}_{1,1}+2\hat{d}\hat{c}\hat{\gamma}_2,\end{cases}$$

其中 $\hat{d}=-\hat{a}+\hat{b}\hat{\mu}_n$,求解得到

$$\hat{\alpha}_{(1,1)}=\frac{\hat{\gamma}_{3,3}\hat{\gamma}_1-\hat{\gamma}_{2,2}\hat{\gamma}_2}{\hat{\gamma}_{2,2}\hat{\gamma}_1-\hat{\gamma}_{1,1}\hat{\gamma}_2}.\tag{7.1.56}$$

从而有

$$\begin{cases}\hat{b}=\dfrac{\hat{\alpha}(1,1)-\hat{a}^2}{\hat{c}};\\ \hat{\sigma}^2=\hat{c}/\hat{b}.\end{cases}\tag{7.1.57}$$

现在讨论双重线性模型 AR(1)-MA(0)参数的矩估计,即

$$x_t=(\theta+\eta_t)x_{t-1}+\varepsilon_t,$$

其中$\{\varepsilon_t\}$、$\{\eta_t\}$及 θ 满足例 7.1.4 的条件, 记 $\gamma_k^*=Ex_tx_{t+k}$, $R_k^*=Ex_t^2x_{t+k}^2$. 容易验证

$$\begin{cases}\gamma_k^*=\theta\gamma_{k-1}^*,\ k\geqslant 1;\\ R_k^*=(\theta^2+\sigma_\eta^2)R_{k-1}^*+\sigma^2\gamma_0^*.\end{cases}\tag{7.1.58}$$

在(7.1.58)式的第一个方程中取 $k=1$,第二个方程中取 $k=1,2$,并用样本矩 $\hat{\gamma}_k^*$ 和 $\hat{R}_k^*$ 代替 γ_k^* 和 R_k^*,得到

$$\begin{cases}\hat{\gamma}_1^*=\theta\hat{\gamma}_0^*;\\ \hat{R}_1^*=(\theta^2+\sigma_\eta^2)\hat{R}_0^*+\sigma_\varepsilon^2\hat{\gamma}_0^*;\\ \hat{R}_2^*=(\theta^2+\sigma_\eta^2)\hat{R}_1^*+\sigma_\varepsilon^2\hat{\gamma}_0^*,\end{cases}\tag{7.1.59}$$

其中

$$\hat{\gamma}_k^*=\frac{1}{n}\sum_{t=1}^{n-k}x_tx_{t+k}\quad(k=0,1,\cdots,n-1),$$

$$\hat{R}_k^* = \frac{1}{n}\sum_{t=1}^{n-k} x_t^2 x_{t+k}^2 \quad (k = 0,1,\cdots,n-1).$$

求解方程组(7.1.59)得

$$\begin{cases} \hat{\theta} = \hat{\gamma}_1^* / \hat{\gamma}_0^*, \\ \hat{\sigma}_\eta^2 = [(\hat{R}_2^* - \hat{R}_1^*)/(\hat{R}_1^* - \hat{R}_0^*)] - \hat{\theta}^2, \\ \hat{\sigma}_\varepsilon^2 = [\hat{R}_1^* - (\hat{\theta}^2 + \hat{\sigma}_\eta^2)\hat{R}_0^*]/\hat{\gamma}_0^*. \end{cases}$$

对更一般的双重线性模型 AR(1)-MA(q),$q \geqslant 2$,其相关结构可由模型的二阶矩和四阶矩来刻画,但此时表达式十分复杂,读者可参阅文献[124].

最后我们指出,对于有限参数型的模型,如果它有平稳解和适当阶矩时,一般说来,都能导出矩与参数的关系式.从而当以样本矩代替相应的理论值时,可导出参数的矩估计.显然,这样的估计的精度较差.但是矩估计方法易于理解,而且依大样本理论,易得到其相容性质.所以可将其作为计算其他估计方法的迭代初值.在上述的矩与参数的关系中,不一定局限于只考虑普通 的整数阶矩,也可以包括局部矩.现以如下简单的 TAR 模型为例说明之,即

$$x_t = \begin{cases} \alpha x_{t-1} + \varepsilon_t, & \text{当 } x_{t-1} \geqslant 0; \\ \beta x_{t-1} + \varepsilon_t, & \text{当 } x_{t-1} < 0, \end{cases} \tag{7.1.60}$$

其中$\{\varepsilon_t\}$是白噪声序列,满足(3.2.2)式,且系数 α、β 满足 $|\alpha| + |\beta| < 1$,则由 § 4.3 和 § 4.5 的讨论知,模型(7.1.60)有唯一的平稳解和二阶矩.为了导出模型(7.1.60)中的参数与矩的关系,我们把(7.1.60)式改写为

$$x_t = \alpha I(x_{t-1} \geqslant 0) x_{t-1} + \beta I(x_{t-1} < 0) x_{t-1} + \varepsilon_t. \tag{7.1.61}$$

由(7.1.61)容易算得

$$\begin{cases} Ex_t = \alpha E x_{t-1} I(x_{t-1} \geqslant 0) \\ \qquad\qquad + \beta E x_{t-1} I(x_{t-1} < 0); \\ Ex_t x_{t-1} = \alpha E x_{t-1}^2 I(x_{t-1} \geqslant 0) \\ \qquad\qquad + \beta E x_{t-1}^2 I(x_{t-1} < 0); \\ Ex_t^2 = \alpha^2 E x_{t-1}^2 I(x_{t-1} \geqslant 0) \\ \qquad\qquad + \beta^2 E x_{t-1}^2 I(x_{t-1} < 0) + \sigma^2. \end{cases} \tag{7.1.62}$$

若在(7.1.62)式中,将 $Ex_t, Ex_t x_{t-1}, Ex_t^2, Ex_{t-1}I(x_{t-1} \geqslant 0), Ex_{t-1}I(x_{t-1} < 0), Ex_{t-1}^2 I(x_{t-1} \geqslant 0)$和 $Ex_{t-1}^2 I(x_{t-1} < 0)$用样本值代替之,便可求参数 α、β 和 σ^2 的矩估计,其中 $Ex_t I(x_{t-1} \geqslant 0)$、$Ex_{t-1}^2 I(x_{t-1} \geqslant 0)$等就是所谓的局部矩.

三、极大似然估计

设 $x_1,\cdots,x_n$ 是来自平稳时间序列 $\{x_t,t\geqslant 1\}$ 的样本，$P_n(x_1,\cdots,x_n)$ 是 $\boldsymbol{x}_n=(x_1,\cdots,x_n)^\tau$ 的概率分布(或密度)函数，它依赖于参数 $\theta\in\Theta$，Θ 为 R^p 中一个开集. 为了简便，我们暂时考虑 $p=1$ 的情形. 有时为强调 $\boldsymbol{x}_n$ 的概率分布对参数的依赖，也记其概率分布(或密度)函数为 $P_n(x_1,\cdots,x_n;\theta)$. 令

$$L_n(\theta)=\log P_n(x_1,\cdots,x_n;\theta)$$

是 $\boldsymbol{x}_n$ 的对数似然函数，并记

$$u_i(\theta)=\mathrm{d}[L_i(\theta)-L_{i-1}(\theta)]/\mathrm{d}\theta,$$
$$L_0(\theta)=0.$$

假定下列条件成立：

(i) $L_n(\theta)$ 关于 θ 是可微的；

(ii) 对每个 $n\geqslant 1$，$E_\theta\left[\frac{\mathrm{d}}{\mathrm{d}\theta}L_n(\theta)\right]^2<\infty$；

(iii) $\int P_n(x_1,\cdots,x_n;\theta)\mathrm{d}\boldsymbol{x}_n$ 在积分号下关于 θ 是两次可微的.

根据 $u_i(\theta)$ 和 $L_n(\theta)$ 的定义知

$$\frac{\mathrm{d}L_n(\theta)}{\mathrm{d}\theta}=\sum_{i=1}^n u_i(\theta), \tag{7.1.63}$$

且有如下的 Taylor 展开式

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\theta}L_n(\theta')&=\sum_{i=1}^n u_i(\theta')\\&=\sum_{i=1}^n u_i(\theta)+(\theta'-\theta)I_n(\theta)\\&\quad+(\theta'-\theta)(J_n(\theta_n^*)+I_n(\theta))\end{aligned} \tag{7.1.64}$$

其中 $\theta'\in\Theta$，$\theta_n^*=\theta+\gamma(\theta'-\theta)$，$\gamma=\gamma(\boldsymbol{x}_n,\theta)$，且 $|\gamma|<1$.

$$I_n(\theta)=\sum_{i=1}^n E_\theta\{u_i^2(\theta)\,|\,\mathscr{F}_{1,i-1}^x\},$$

$$J_n(\theta)=\sum_{i=1}^n\frac{\mathrm{d}u_i(\theta)}{\mathrm{d}\theta}.$$

由于 $E_\theta\{u_i(\theta)\,|\,\mathscr{F}_{1,i-1}^x\}=0$，a.s.，故 $\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}L_n(\theta),\mathscr{F}_{1,n}^x,n\geqslant 1\right\}$ 是平方可积鞅. 假定 $I_n(\theta)\to\infty$，a.s. (当 $n\to\infty$ 时)，由

$$\sum_{n=1}^{\infty}[I_n(\theta)]^2E_\theta\{u_n^2(\theta)\,|\,\mathscr{F}_{1,n-1}^x\}<\infty,\ \text{a.s.},$$

及鞅的强大数定律，有

$$[I_n(\theta)]^{-1}\sum_{i=1}^{n}u_i(\theta)\to 0,\ \text{a.s.},\ 当\ n\to\infty. \tag{7.1.65}$$

令 $\hat{\theta}_n$ 是似然方程

$$\frac{\mathrm{d}}{\mathrm{d}\theta}L_n(\theta)=0$$

的解，则由(7.1.64)式有

$$[I_n(\theta)]^{-1}\sum_{i=1}^{n}u_i(\theta)=(\hat{\theta}_n-\theta)\left[1-\frac{J_n(\theta_n^*)+I_n(\theta)}{I_n(\theta)}\right]. \tag{7.1.66}$$

如果再假定

$$\limsup_{n\to\infty}[I_n(\theta)]^{-1}|J_n(\theta_n^*)+I_n(\theta)|<1,\ \text{a.s.},$$

则根据(7.1.65)和(7.1.66)式知，

$$\hat{\theta}_n\to\theta,\ \text{a.s.}\quad(当\ n\to\infty), \tag{7.1.67}$$

即 $\hat{\theta}_n$ 是真值 θ 的强相容估计.

由(7.1.66)式可见，如果

$$[I_n(\theta)]^{1/2}\sum_{i=1}^{n}u_i(\theta)\xrightarrow{d}N(0,1),$$

且在 Θ 的一个包含真值 θ 的紧子集上，一致地有

$$\frac{J_n(\theta)}{I_n(\theta)}\xrightarrow{p}-1\quad(当\ n\to\infty),$$

这里 $\xrightarrow{p}$ 表示依概率收敛，则

$$[I_n(\theta)]^{1/2}(\hat{\theta}_n-\theta)\xrightarrow{d}N(0,1).$$

于是我们有如下定理，其证明参阅文献[56]、[106].

定理 7.1.9 如果

(i) $I_n(\theta)\to\infty$, a.s. (当 $n\to\infty$),

$$\frac{I_n(\theta)}{EI_n(\theta)}\xrightarrow{p}\eta\quad(\neq 0),$$

其中 η 是一个随机变量，且在 Θ 中一个包含真值 θ 的紧子集上

$$\frac{J_n(\theta)}{I_n(\theta)}\xrightarrow{p}-1;$$

(ii) 对 $\delta>0$, $|\theta_n-\theta|<\delta/(E_\theta I_n(\theta))^{1/2}$,

$$E_{\theta_n} I_n(\theta_n) = E_\theta I_n(\theta)(1 + 0(1)),$$

$$I_n(\theta_n) = I_n(\theta)(1 + o(1)), \text{a.s.},$$

$$J_n(\theta_n) = J_n(\theta) + o(I_n(\theta)), \text{a.s.},$$

则
$$[I_n(\theta)]^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0,1) \quad (当\ n \to \infty).$$

以下讨论非线性时间序列模型的极大似然估计.

定理 7.1.10　考虑非线性模型(7.1.14),即

$$x_t = \varphi(x_{t-1};\theta) + \varepsilon_t, \tag{7.1.68}$$

其中 $\varphi(\cdot;\theta)$ 关于 θ 有二阶连续导数, $E_\theta[\mathrm{d}\varphi(x_{t-1};\theta)/\mathrm{d}\theta]^2>0$, $\{\varepsilon_t\}$ 是正态白噪声序列, $\varepsilon_t \sim N(0,1)$. $\boldsymbol{x}_n=(x_1,\cdots,x_n)^\tau$ 是来自模型(7.1.68)的样本,且 $\{\varepsilon_1,\cdots,\varepsilon_n\}$ 与 x_0 独立. 容易验证

$$\begin{aligned} &P_n(x_1,\cdots,x_n;\theta|x_0) \\ &= \prod_{t=1}^{n} P_1(x_t;\theta|x_{t-1}) \\ &= (2\pi)^{-n/2}\prod_{t=1}^{n}\exp\left\{-\frac{1}{2}[x_t - \varphi(x_{t-1};\theta)]^2\right\}. \end{aligned}$$

从而对数似然函数为

$$L_n(\theta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}[x_t - \varphi(x_{t-1};\theta)]^2.$$

由

$$\frac{\mathrm{d}L_n(\theta)}{\mathrm{d}\theta} = 0$$

得到对数似然方程

$$\begin{aligned} \frac{\mathrm{d}L_n(\theta)}{\mathrm{d}\theta} &= \sum_{t=1}^{n}[x_t - \varphi(x_{t-1};\theta)]\frac{\mathrm{d}\varphi(x_{t-1};\theta)}{\mathrm{d}\theta} \\ &= 0, \end{aligned} \tag{7.1.69}$$

记其解为 $\hat{\theta}_n$. 令 $L_0(\theta)=0$,由(7.1.63)式和(7.1.69)式,有

$$\begin{aligned} u_i(\theta) &= \frac{\mathrm{d}}{\mathrm{d}\theta}[L_i(\theta) - L_{i-1}(\theta)] \\ &= \left[\frac{\mathrm{d}}{\mathrm{d}\theta}\varphi(x_{i-1};\theta)\right]\varepsilon_i \quad (i=1,\cdots,n). \end{aligned}$$

故

$$\begin{aligned} E_\theta\{u_i(\theta)|\mathscr{F}_{1i-1}^{\varepsilon}\} &= \frac{\mathrm{d}}{\mathrm{d}\theta}\varphi(x_{i-1};\theta)E\varepsilon_i \\ &= 0, \text{ a.s.} \end{aligned} \tag{7.1.70}$$

$$I_n(\theta)=\sum_{i=1}^{n}E_\theta\{u_i^2(\theta)\,|\,\mathscr{F}_{1,i-1}^{\tau}\}$$

$$=\sum_{i=1}^{n}\left[\frac{\mathrm{d}}{\mathrm{d}\theta}\varphi(x_{i-1};\theta)\right]^2, \tag{7.1.71}$$

$$J_n(\theta)=\sum_{i=1}^{n}\frac{\mathrm{d}u_i(\theta)}{\mathrm{d}\theta}$$

$$=\sum_{i=1}^{n}\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\varphi(x_{t-1};\theta)\varepsilon_i-\sum_{i=1}^{n}\left[\frac{\mathrm{d}}{\mathrm{d}\theta}\varphi(x_{t-1};\theta)\right]^2, \tag{7.1.72}$$

由遍历性，并注意到 $E_\theta\left[\frac{\mathrm{d}}{\mathrm{d}\theta}\varphi(x_{t-1};\theta)\right]^2>0$，可知当 $n\to\infty$ 时，

$$I_n(\theta)\to\infty,\ \text{a.s.},\quad [I_n(\theta)]^{-1}[J_n(\theta_n^*)+I_n(\theta)]\to 0,\ \text{a.s.}.$$

根据(7.1.67)式的讨论知，由似然方程(7.1.69)求得的解 $\hat{\theta}_n$ 是真参数 θ 的强相容估计.

注意到 $\varphi(\cdot;\theta)$ 关于 θ 为二阶连续可微，则由(7.1.70)～(7.1.72)式，容易验证定理7.1.9的条件满足. 因此，由(7.1.69)求得的MLE$\hat{\theta}_n$ 还具有渐近正态性.

例7.1.11 考虑AR(1)-MA(0)模型(7.1.20)，即

$$x_t=(\theta+\eta_t)x_{t-1}+\varepsilon_t.$$

除了例7.1.4的假定外，再假定$\{\varepsilon_t\}$和$\{\eta_t\}$都是正态白噪声序列，不难验证

$$E\{x_t\,|\,x_{t-1}\}=E\{(\theta+\eta_t)x_{t-1}+\varepsilon_t\,|\,x_{t-1}\}$$

$$=\theta x_{t-1},$$

$$E\{x_t^2\,|\,x_{t-1}\}=E\{[(\theta+\eta_t)x_{t-1}+\varepsilon_t]^2\,|\,x_{t-1}\}$$

$$=\theta^2x_{t-1}^2+\sigma_\eta^2x_{t-1}^2+\sigma^2.$$

故

$$\mathrm{Var}(x_t\,|\,x_{t-1})=E\{x_t^2-[E\{x_t\,|\,x_{t-1}\}]^2\,|\,x_{t-1}\}$$

$$=\sigma_\eta^2x_{t-1}^2+\sigma^2.$$

给定样本 $\boldsymbol{x}_n=(x_1,\cdots,x_n)^\tau$，有

$$P_n(x_1,\cdots,x_n\,|\,x_0)=\prod_{t=1}^{n}P_1(x_t\,|\,x_{t-1})$$

$$=(2\pi)^{-n/2}\prod_{t=1}^{n}(\sigma_\eta^2x_{t-1}^2+\sigma^2)^{-1}$$

$$\times\exp\left\{-\frac{1}{2}\frac{(x_t-\theta x_{t-1})^2}{\sigma_\eta^2x_{t-1}^2+\sigma_2^2}\right\},$$

从而对数似然函数为

$$L_n(\theta,\sigma_\eta^2,\sigma_\varepsilon^2)=-\frac{n}{2}\log(2\pi)-\frac{1}{2}\sum_{t=1}^{n}\log(\sigma_\eta^2x_{t-1}^2+\sigma_\varepsilon^2)-\frac{1}{2}\sum_{t=1}^{n}\frac{(x_t-\theta x_{t-1})^2}{\sigma_\eta^2x_{t-1}^2+\sigma_\varepsilon^2}. \tag{7.1.73}$$

显然，$L_n(\theta,\sigma_\eta^2,\sigma_\varepsilon^2)$是$\sigma_\varepsilon^2$和$\sigma_\eta^2$的非线性函数. 令$\lambda=\sigma_\eta^2/\sigma_\varepsilon^2$，则(7.1.73)式变为

$$L_n(\theta,\lambda,\sigma_\varepsilon^2)=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log\sigma_\varepsilon^2-\frac{1}{2}\sum_{t=1}^{n}\log(1+\lambda x_{t-1}^2)-\frac{1}{2\sigma_\varepsilon^2}\sum_{t=1}^{n}\frac{(x_t-\theta x_{t-1})^2}{1+\lambda x_{t-1}^2}. \tag{7.1.74}$$

令

$$\bar{l}_n(\theta,\lambda,\sigma_\varepsilon^2)=\log\sigma_\varepsilon^2+\frac{1}{n}\sum_{t=1}^{n}\log(1+\lambda x_{t-1}^2)+\frac{1}{n\sigma_\varepsilon^2}\sum_{t=1}^{n}\frac{(x_t-\theta x_{t-1})^2}{1+\lambda x_{t-1}^2}. \tag{7.1.75}$$

为求模型(7.1.20)的参数θ、σ_ε^2和σ_η^2的MLE，只需关于θ、σ_ε^2和λ求对数似然函数(7.1.74)的极大值. 与此等价地，只需关于θ、σ_ε^2和λ求(7.1.75)的极小值. 由于

$$\frac{\partial}{\partial\sigma_\varepsilon^2}\bar{l}_n(\theta,\lambda,\sigma_\varepsilon^2)=\sigma_\varepsilon^{-2}-\frac{1}{n\sigma_\varepsilon^4}\sum_{t=1}^{n}\frac{(x_t-\theta x_{t-1})^2}{1+\lambda x_{t-1}^2},$$

故由

$$\frac{\partial}{\partial\sigma_\varepsilon^2}\bar{l}_n(\theta,\lambda,\sigma_\varepsilon^2)=0,$$

解得

$$\sigma_{\varepsilon n}^2(\lambda,\theta)=\frac{1}{n}\sum_{t=1}^{n}\frac{(x_t-\theta x_{t-1})^2}{1+\lambda x_{t-1}^2}. \tag{7.1.76}$$

再令

$$\tilde{l}_n(\theta,\lambda)=\inf_{\sigma^2>0}\bar{l}_n(\theta,\lambda,\sigma_\varepsilon^2)-1, \tag{7.1.77}$$

则由(7.1.75)和(7.1.76)两式知(7.1.77)式又可表示为

$$\tilde{l}_n(\theta,\lambda)=\frac{1}{n}\sum_{t=1}^{n}\log(1+\lambda x_{t-1}^2)$$
$$+\log\left\{\frac{1}{n}\sum_{t=1}^{n}\frac{(x_t-\theta x_{t-1})^2}{1+\lambda x_{t-1}^2}\right\}. \tag{7.1.78}$$

关于 θ 和 γ 极小化 $\tilde{l}_n(\theta,\lambda)$，得到 θ 和 λ 的 MLE $\hat{\theta}_n$ 和 $\hat{\lambda}_n$，即

$$\tilde{l}_n(\hat{\theta}_n,\hat{\lambda}_n)=\inf_{(\theta,\lambda)^\tau\in\Theta'}\tilde{l}_n(\theta,\lambda),$$

其中 $\Theta'\subset R^2$ 为一开集. 再根据(7.1.76)式得到 $\hat{\sigma}_\varepsilon^2$ 的 MLE：

$$\hat{\sigma}_{\varepsilon n}^2=\frac{1}{n}\sum_{t=1}^{n}\frac{(x_t-\hat{\theta}_n x_{t-1})^2}{1+\hat{\lambda}_n x_{t-1}^2},$$

$$\hat{\sigma}_\eta^2=\hat{\sigma}_{\varepsilon n}^2\hat{\lambda}_n.$$

下面讨论比较一般的双重线性模型参数的极大似然估计. 为此仍考虑双重线性模型的第一重模型为

$$x_t=\varphi_t x_{t-1}+\varepsilon_t\quad(t\geqslant 1). \tag{7.1.79}$$

假定 $\{\varphi_t\}$ 与 $\{\varepsilon_t\}$ 为相互独立的随机序列，且初值 x_0 与 $\{\varepsilon_s,s\geqslant 1\}$ 独立. $\{\varepsilon_t\}$ 还是正态白噪声序列，$\varepsilon_t\sim N(0,\sigma^2)$. 至于(7.1.79)式中 $\{\varphi_t\}$ 所满足的有限参数模型，在此不具体写出，而用更广泛的形式描述，即假定它的有限维分布被有限维参数向量 $\boldsymbol{\theta}$ 所决定，$\boldsymbol{\theta}\in\Theta$，$\Theta\subset R^p$ 为一个开集. 记

$$\boldsymbol{x}_n=(x_1,\cdots,x_n)^\tau,$$
$$\boldsymbol{\varepsilon}^{(n)}=(\varepsilon_1,\cdots,\varepsilon_n)^\tau,$$
$$\boldsymbol{\varphi}_\theta^{(n)}=(\varphi_1^{(\theta)},\cdots,\varphi_n^{(\theta)})^\tau.$$

利用以上记号和假定，在给定 $\boldsymbol{\varphi}_\theta^{(n)}$ 的条件下，$\boldsymbol{\varepsilon}^{(n)}$ 的条件密度为

$$f_{\theta,\sigma^2}(\boldsymbol{\varepsilon}^{(n)}\mid\boldsymbol{\varphi}_\theta^{(n)})=u(\varepsilon_1,\cdots,\varepsilon_n,\sigma^2), \tag{7.1.80}$$

式中 u 表示 n 维正态密度函数，其均值为$\mathbf{0}$(零向量)，方差为 $\sigma^2 I_n$，I_n 为 n 阶单位矩阵. 再由(7.1.79)式和(7.1.80)式可知，在给定 $\boldsymbol{\varphi}_\theta^{(n)}$ 和 x_0 的条件下，$\boldsymbol{x}_n$ 的条件密度为

$$g_{\theta,\sigma^2}(\boldsymbol{x}_n\mid\boldsymbol{\varphi}_\theta^{(n)},x_0)$$
$$=u(x_1-\varphi_1^{(\theta)}x_0,x_2-\varphi_2^{(\theta)}x_1,\cdots,x_n-\varphi_n^{(\theta)}x_{n-1},\sigma^2). \tag{7.1.81}$$

最后，再由(7.1.81)式可知，在给定 x_0 的条件下，$\boldsymbol{x}_n$ 的条件密度为

$$h_{\theta,\sigma^2}(\boldsymbol{x}_n\mid x_0)=\int_{R^n}g_{\theta,\sigma^2}(\boldsymbol{x}_n\mid\boldsymbol{\varphi}_\theta^{(n)},x_0)\mathrm{d}F(\boldsymbol{\varphi}_\theta^{(n)}), \tag{7.1.82}$$

式中 $F(\boldsymbol{\varphi}_\theta^{(n)})$ 为给定 $\boldsymbol{\theta}\in\Theta$ 时 $\boldsymbol{\varphi}_\theta^{(n)}$ 的分布函数.

对给定的样本 $\boldsymbol{x}_n$，用(7.1.82)式求出 $\boldsymbol{x}_n$ 的对数条件似然函数

$$L_n(\theta,\sigma^2) = \log h_{\theta,\sigma^2}(\boldsymbol{x}_n | x_0). \tag{7.1.83}$$

再关于 θ、σ^2求(7.1.83)式的极大值解 $\hat{\theta}_n$、$\hat{\sigma}_n^2$，即

$$L_n(\hat{\theta}_n,\hat{\sigma}_n^2) = \sup_{\theta\in\Theta,\sigma^2>0} L_n(\theta,\sigma^2).$$

所求的 $\hat{\theta}_n$、$\hat{\sigma}_n^2$ 即为 θ 和 σ^2的 MLE. 但是，由于以上公式的计算十分复杂，很难求解 θ 和 σ^2，一般可用以下近似解法.

为了求解(7.1.83)式中 $L_n(\theta,\sigma^2)$的极大值，首先给出计算 $L_n(\theta,\sigma^2)$的一种实用算法. 这种算法是将(7.1.82)式右端对$g_{\theta,\sigma^2}(\boldsymbol{x}_n|\boldsymbol{\varphi}_\theta^{(n)},x_0)$的依概率分布 $F(\boldsymbol{\varphi}_\theta^{(n)})$的平均，用其样本的平均代替. 即用蒙特卡洛方法模拟产生$\boldsymbol{\varphi}_\theta^{(n)}$的 k 个相互独立的样本 $\boldsymbol{\varphi}_{\theta 1}^{(n)},\cdots,\boldsymbol{\varphi}_{\theta k}^{(n)}$，并使这 k 个相互独立的样本具有共同的分布 $F(\boldsymbol{\varphi}_\theta^{(n)})$，然后用如下样本平均代替(7.1.82)式右端的积分：

$$\hat{h}_{\theta,\sigma^2}(\boldsymbol{x}_n | x_0) = \frac{1}{k}\sum_{i=1}^{k} g_{\theta,\sigma^2}(\boldsymbol{x}_n | \boldsymbol{\varphi}_{\theta i}^{(n)},x_0). \tag{7.1.84}$$

式中样本个数 k 可视精度要求而定.

以上述求得的 $\hat{h}_{\theta,\sigma^2}(\boldsymbol{x}_n|x_0)$作为 $h_{\theta,\sigma^2}(\boldsymbol{x}_n|x_0)$的近似值并代替(7.1.83)式中的 $h_{\theta,\sigma^2}(\boldsymbol{x}_n|x_0)$，然后将依此计算出来的 $\hat{L}_n(\theta,\sigma^2)$作为 $L_n(\theta,\sigma^2)$的近似值. 对所有可能的参数 θ、σ^2，求 $\hat{L}_n(\theta,\sigma^2)$的极大值，以此作为 θ 和 σ^2的 MLE. 具体步骤从略. 下面以模型(7.1.20)说明之. 记 $\boldsymbol{\theta}=(\theta,\sigma_\eta^2)$，根据(7.1.82)式可知，在给定 $\boldsymbol{\varphi}_\theta^{(n)}$的条件下，$\varepsilon^{(n)}$的条件密度为

$$\begin{aligned} f_{\theta,\sigma^2}(\varepsilon^{(n)} | \boldsymbol{\varphi}_\theta^{(n)}) &= u(\varepsilon_1,\cdots,\varepsilon_n,\sigma^2)) \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{n}\varepsilon_t^2\right\}, \end{aligned} \tag{7.1.85}$$

式中 $\varphi_t=\theta+\eta_t$. 从而由(7.1.82)式可得

$$\begin{aligned} &g_{\theta,\sigma^2}(\boldsymbol{x}_n | \boldsymbol{\varphi}_\theta^{(n)},x_0) \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j-\varphi_j x_{j-1})^2\right\}. \end{aligned} \tag{7.1.86}$$

再注意到 $\varphi_t=\theta+\eta_t$，$\{\eta_t\}$是独立同分布的正态随机序列， $\eta_t\sim N(0,\sigma_\eta^2)$，可知 $F(\boldsymbol{\varphi}_\theta^{(n)})$为 n 维正态分布，其联合密度函数为

$$f_{\theta,\sigma^2}(\boldsymbol{\varphi}_\theta^{(n)}) = (2\pi\sigma_\eta^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma_\eta^2}\sum_{t=1}^{n}(\varphi_t-\theta)^2\right\}. \tag{7.1.87}$$

于是再用(7.1.82)式可得

$$h_{\theta,\sigma^2}(\boldsymbol{x}_n | x_0)$$

$$= \int_{R^n} (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^n (x_j - \varphi_j x_{t-j})^2\right\}$$

$$\times (2\pi\sigma_\eta^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_\eta^2}\sum_{j=1}^n (\varphi_j - \theta)^2\right\} \mathrm{d}\varphi_1, \cdots, \mathrm{d}\varphi_n$$

$$= (2\pi\sigma\sigma_\eta)^{-n} \int_{R^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^n [(x_j - \theta x_{j-1}) - \eta_j x_{j-1}]^2\right\}$$

$$\times \exp\left\{-\frac{1}{2\sigma_\eta^2}\sum_{j=1}^n \eta_j^2\right\} \mathrm{d}\eta_1, \cdots, \mathrm{d}\eta_n. \qquad (7.1.88)$$

对(7.1.83)式右端积分,经过繁琐的计算后,不难验证

$$h_{\theta,\sigma^2}(\boldsymbol{x}_n | x_0) = (2\pi)^{-\frac{n}{2}} \prod_{j=1}^n (\sigma^2 + \sigma_\eta^2 x_{j-1}^2)^{-1}$$

$$\times \exp\left\{-\frac{1}{2}\frac{(x_j - \theta x_{j-1})^2}{\sigma^2 + \sigma_\eta^2 x_{j-1}^2}\right\}. \qquad (7.1.89)$$

对(7.1.89)式两端取对数,得 $\boldsymbol{x}_n$ 的对数条件似然函数

$$L_n(\boldsymbol{\theta}, \sigma^2) = \log h_{\theta,\sigma^2}(\boldsymbol{x}_n | x_0)$$

$$= -\frac{n}{2}\log(2\pi) - \sum_{j=1}^n \log(\sigma^2 + \sigma_\eta^2 x_{j-1}^2)$$

$$- \frac{1}{2}\sum_{j=1}^n \frac{(x_j - \theta x_{j-1})^2}{\sigma^2 + \sigma_\eta^2 x_{j-1}^2}.$$

上式与(7.1.74)相同. 按(7.1.74)式以下讨论的估计方法可直接求出 MLE,而不必使用前述的蒙特卡洛方法求其近似值.

在一般情况下,(7.1.83)式的 $L_n(\boldsymbol{\theta}, \sigma^2)$ 是无法直接计算的,而只能用近似方法计算. 所以其解也只能是近似解. 除了前面介绍的蒙特卡洛近似方法外,寻找其他更有效的近似方法也是有实用价值的研究工作.

现在讨论 AR (p) -ARCH(q) 模型参数的极大似然估计,考虑模型(6.5.23),即

$$\begin{cases} x_t = \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + e_t; \\ e_t | \mathscr{F}_{t-1}^x \sim N(0, h_t); \\ h_t = \varphi_0 + \varphi_1 e_{t-1}^2 + \cdots + \varphi_q e_{t-q}^2. \end{cases} \qquad (7.1.90)$$

对模型(7.1.90)的一组样本 $\boldsymbol{x}_n = (x_1, \cdots, x_n)^\tau$,由(6.5.24)和(6.5.25)两式知,其关于$(x_0, x_{-1}, \cdots, x_{-p+1})$的条件似然函数为

$$L(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^n \log f(x_t | x_{t-1}, \cdots, x_{t-p})$$

$$= \frac{1}{n}\sum_{t=1}^{n}\left(-\frac{1}{2}\log h_t - \frac{e_t^2}{2h_t}\right), \qquad (7.1.91)$$

式中 $\boldsymbol{\theta}=(\boldsymbol{\alpha}^\tau,\boldsymbol{\varphi}^\tau)^\tau$, $\boldsymbol{\alpha}=(\alpha_1,\cdots,\alpha_p)^\tau$, $\boldsymbol{\varphi}=(\varphi_0,\varphi_1,\cdots,\varphi_q)^\tau$. 令

$$l_t = -\frac{1}{2}\log h_t - \frac{e_t^2}{2h_t},$$

则 l_t 关于 $\boldsymbol{\alpha}=(\alpha_1,\cdots,\alpha_p)^\tau$ 的一阶和二阶偏导数为

$$\frac{\partial l_t}{\partial \boldsymbol{\alpha}} = \frac{e_t \boldsymbol{x}_{t-1}}{h_t} + \frac{1}{2h_t}\left(\frac{\partial h_t}{\partial \boldsymbol{\alpha}}\right)\left(\frac{e_t^2}{h_t} - 1\right),$$

$$\frac{\partial^2 l_t}{\partial \boldsymbol{\alpha}\partial \boldsymbol{\alpha}^\tau} = -\frac{\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}^\tau}{h_t} - \frac{1}{2h_t}\cdot\frac{\partial h_t}{\partial \boldsymbol{\alpha}}\cdot\frac{\partial h_t}{\partial \boldsymbol{\alpha}^\tau}\left(\frac{e_t^2}{h_t}\right)$$

$$- \frac{2e_t\boldsymbol{x}_{t-1}}{h_t^2}\left(\frac{\partial h_t}{\partial \boldsymbol{\alpha}^\tau}\right) + \left(\frac{e_t^2}{h_t} - 1\right)\frac{\partial}{\partial \boldsymbol{\alpha}^\tau}\left(\frac{1}{2h_t}\frac{\partial h_t}{\partial \boldsymbol{\alpha}}\right).$$

其中 $\boldsymbol{x}_t=(x_t,\cdots,x_{t-p+1})^\tau$. 注意到

$$\frac{\partial h_t}{\partial \boldsymbol{\alpha}} = \frac{\partial}{\partial \boldsymbol{\alpha}}\left[\varphi_0 + \sum_{j=1}^{q}\varphi_j(x_{t-j} - \boldsymbol{\alpha}^\tau \boldsymbol{x}_{t-1-j})^2\right]$$

$$= -2\sum_{j=1}^{q}\varphi_j e_{t-j}\boldsymbol{x}_{t-1-j},$$

则有

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}} = \frac{1}{n}\sum_{t=1}^{n}\left[\frac{e_t\boldsymbol{x}_{t-1}}{h_t} - \frac{1}{h_t}\left(\frac{e_t^2}{h_t} - 1\right)\sum_{j=1}^{q}\varphi_j e_{t-j}\boldsymbol{x}_{t-1-j}\right],$$

$$I_{\alpha\alpha} = \frac{1}{n}\sum_{t=1}^{n}E\left(-\frac{\partial^2 l_t}{\partial \boldsymbol{\alpha}\partial \boldsymbol{\alpha}^\tau}\right)$$

$$= \frac{1}{n}\sum_{t=1}^{n}E\left\{E\left\{-\frac{\partial^2 l_t}{\partial \boldsymbol{\alpha}\partial \boldsymbol{\alpha}^\tau}\middle|\mathscr{F}_{t-1}^x\right\}\right\}$$

$$= \frac{1}{n}\sum_{t=1}^{n}E\left(\frac{\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}}{h_t} + 2\sum_{j=1}^{q}\varphi_j^2\frac{e_{t-j}^2}{h_t^2}\boldsymbol{x}_{t-1-j}\boldsymbol{x}_{t-1-j}^\tau\right). \qquad (7.1.92)$$

由(6.5.26)式以下的推导知

$$I_{\alpha\varphi} = 0,$$

$$I_{\varphi\varphi} = \frac{1}{n}\sum_{t=1}^{n}E\left(\frac{1}{2h_t}\boldsymbol{z}_t\boldsymbol{z}_t^\tau\right), \qquad (7.1.93)$$

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi}} = \frac{1}{n}\sum_{t=1}^{n}\frac{1}{2h_t}\left(\frac{e_t^2}{h_t} - 1\right)\boldsymbol{z}_t,$$

式中 $\boldsymbol{z}_t=(1,e_{t-1}^2,\cdots,e_{t-q}^2)^\tau$. 这样, $\boldsymbol{\theta}$ 的 MLE $\hat{\boldsymbol{\theta}}=(\hat{\boldsymbol{\alpha}}^\tau,\hat{\boldsymbol{\varphi}}^\tau)^\tau$ 是方程

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}} = 0 \text{ 和 } \frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi}} = 0$$

的解.求上述方程的解只能用数值迭代法.比较常用的方法是 Newton-Raphson 方法.为了方便,记

$$\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \cdots, \hat{\theta}_{p+q+1})^\tau,$$

当初始值 $\boldsymbol{\theta}^0 = (\theta_1^0, \cdots, \theta_{p+q+1}^0)^\tau$ 与 $\hat{\theta}$ 比较接近时,则有

$$\begin{aligned} 0 &= \left.\frac{\partial L(\boldsymbol{\theta})}{\partial \theta_i}\right|_{\theta=\hat{\theta}} \\ &\approx \left.\frac{\partial L(\boldsymbol{\theta})}{\partial \theta_i}\right|_{\theta=\theta^0} + \sum_{j=1}^{p+q+1} \left.\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \theta_i \theta_j}\right|_{\theta=\theta^0} (\hat{\theta}_j - \theta_j^0) \\ &\quad (i = 1, \cdots, p+q+1). \end{aligned} \tag{7.1.94}$$

记

$$\boldsymbol{a}(\boldsymbol{\theta}) = \left(\frac{\partial L(\boldsymbol{\theta})}{\partial \theta_1}, \cdots, \frac{\partial L(\boldsymbol{\theta})}{\partial \theta_{p+q+1}}\right)^\tau,$$

则由(7.1.94)式有

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^0 + I(\boldsymbol{\theta}^0)^{-1}\boldsymbol{a}(\boldsymbol{\theta}^0). \tag{7.1.95}$$

由于

$$I(\boldsymbol{\theta}) = \begin{bmatrix} I_{\alpha\alpha} & 0 \\ 0 & I_{\varphi\varphi} \end{bmatrix},$$

故有

$$I(\boldsymbol{\theta})^{-1} = \begin{bmatrix} I_{\alpha\alpha}^{-1} & 0 \\ 0 & I_{\varphi\varphi}^{-1} \end{bmatrix}.$$

于是由(7.1.95)式又有

$$\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}^0 + I_{\alpha\alpha}^{-1}\boldsymbol{a}(\boldsymbol{\alpha}^0), \tag{7.1.96}$$

$$\hat{\boldsymbol{\varphi}} = \boldsymbol{\varphi}^0 + I_{\varphi\varphi}^{-1}\boldsymbol{a}(\boldsymbol{\varphi}^0), \tag{7.1.97}$$

式中

$$\boldsymbol{a}(\boldsymbol{\alpha}) = \left(\frac{\partial L(\boldsymbol{\theta})}{\partial \alpha_1}, \cdots, \frac{\partial L(\boldsymbol{\theta})}{\partial \alpha_p}\right)^\tau,$$

$$\boldsymbol{a}(\boldsymbol{\varphi}) = \left(\frac{\partial L(\boldsymbol{\theta})}{\partial \varphi_0}, \frac{\partial L(\boldsymbol{\theta})}{\partial \varphi_1}, \cdots, \frac{\partial L(\boldsymbol{\theta})}{\partial \varphi_q}\right)^\tau,$$

$\boldsymbol{\alpha}^0$、$\boldsymbol{\varphi}^0$ 为初始值.由(7.1.96)和(7.1.97)两式得到如下迭代公式:

$$\hat{\boldsymbol{\alpha}}^{(k+1)} = \hat{\boldsymbol{\alpha}}^{(k)} + (I_{\alpha\alpha}^{(k)})^{-1}\boldsymbol{a}(\hat{\boldsymbol{\alpha}}^{(k)}) \quad (k = 0, 1, 2, \cdots), \tag{7.1.98}$$

$$\hat{\boldsymbol{\varphi}}^{(k+1)} = \hat{\boldsymbol{\varphi}}^{(k)} + (I_{\varphi\varphi}^{(k)})^{-1}\boldsymbol{a}(\hat{\boldsymbol{\varphi}}^{(k)}) \quad (k = 0,1,2,\cdots). \tag{7.1.99}$$

根据上述讨论,可归纳如下建模步骤:

第一步:根据样本 $x_1,\cdots,x_n$,利用 LAR(p)模型熟知的参数估计方法求模型

$$x_t = \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + e_t$$

的参数 $\boldsymbol{\alpha}$ 的估计,记为 $\hat{\boldsymbol{\alpha}}^{(0)}$. 令

$$\hat{e}_t^{(0)} = x_t - \hat{\alpha}_1^{(0)} x_{t-1} - \cdots - \hat{\alpha}_p^{(0)} x_{t-p} \quad (t = 1,\cdots,n). \tag{7.1.100}$$

第二步:由于对任意的 $k \geqslant 0$,有

$$\begin{aligned} Ee_t^2 e_{t-k}^2 &= E\{E\{e_t^2 e_{t-k}^2 \mid \mathscr{F}_{t-1}^{\tau}\}\} \\ &= E\{e_{t-k}^2 E\{e_t^2 \mid \mathscr{F}_{t-1}^{\tau}\}\} \\ &= E\{e_{t-k}^2(\varphi_0 + \varphi_1 e_{t-1}^2 + \cdots + \varphi_q e_{t-q}^2)\} \\ &= \varphi_1 Ee_{t-1}^2 e_{t-k}^2 + \cdots + \varphi_q Ee_{t-q}^2 e_{t-k}^2 + \varphi_0 Ee_{t-k}^2. \end{aligned} \tag{7.1.101}$$

记 $\gamma_e(0) = Ee_t^2, R_e(k) = Ee_t^2 e_{t-k}^2 \quad (k = 0, \pm 1,\cdots)$,

则由(7.1.101)式有

$$R_e(k) = \varphi_1 R_e(k-1) + \cdots + \varphi_q R_e(k-q) + \varphi_0 r_e(0) \quad (k \geqslant 0). \tag{7.1.102}$$

以估计值

$$\hat{\gamma}_e(0) = \frac{1}{n}\sum_{t=1}^{n}[\hat{e}_t^{(0)}]^2$$

和 $$\hat{R}_e(k) = \frac{1}{n}\sum_{t=k+1}^{n}[\hat{e}_t^{(0)}]^2[\hat{e}_{t-k}^{(0)}]^2 \quad (k = 0,1,\cdots,n-1)$$

代替(7.1.102)式中的 $\gamma_e(0)$和 $R_e(k)$,便可求得 $\boldsymbol{\theta}$ 的矩估计,记为

$$\hat{\varphi}^{(0)} = (\hat{\varphi}_0^{(0)}, \hat{\varphi}_1^{(0)}, \cdots, \hat{\varphi}_p^{(0)})^{\tau}.$$

计算

$$\hat{h}_t^{(0)} = \hat{\varphi}_0^{(0)} + \hat{\varphi}_1^{(0)}[\hat{e}_{t-1}^{(0)}]^2 + \cdots + \hat{\varphi}_q^{(0)}[\hat{e}_{t-q}^{(0)}]^2 \quad (t = q+1,\cdots,n). \tag{7.1.103}$$

第三步:计算

$$\hat{I}_{\alpha\alpha}^{(0)} = \frac{1}{n}\sum_{t=q+1}^{n}\left\{\frac{\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}^{\tau}}{\hat{h}_t^{(0)}} + 2\sum_{j=1}^{q}[\varphi_j^{(0)}]^2 \frac{[\hat{e}_{t-j}^{(0)}]^2}{[\hat{h}_t^{(0)}]^2}\boldsymbol{x}_{t-1-j}\boldsymbol{x}_{t-1-j}^{\tau}\right\}, \tag{7.1.104}$$

$$\hat{I}_{\varphi\varphi}^{(0)} = \frac{1}{n}\sum_{t=q+1}^{n}\frac{1}{2\hat{h}_t^{(0)}}\hat{z}_t\hat{z}_t^{\tau}, \tag{7.1.105}$$

其中 $\hat{z}_t^{(0)}=(1,[\hat{e}_{t-1}^{(0)}]^2,\cdots,[\hat{e}_{t-q}^{(0)}]^2)^{\tau}$.

$$\left.\frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}}\right|_{\theta=\theta^{(0)}} = \frac{1}{n}\sum_{t=q+1}^{n}\left\{\frac{\hat{e}_t^{(0)}\boldsymbol{x}_{t-1}}{\hat{h}_t^{(0)}} - \frac{1}{\hat{h}_t^{(0)}}\left(\frac{[\hat{e}_t^{(0)}]^2}{\hat{h}_t^{(0)}} - 1\right)\right.$$

$$\left.\times\left(\sum_{j=1}^{q}\hat{\varphi}_j^{(0)}\hat{e}_{t-j}^{(0)}\boldsymbol{x}_{t-1-j}\right)\right\}, \tag{7.1.106}$$

$$\left.\frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi}}\right|_{\theta=\theta^{(0)}} = \frac{1}{n}\sum_{t=q+1}^{n}\frac{1}{2\hat{h}_t^{(0)}}\left(\frac{[\hat{e}_t^{(0)}]^2}{\hat{h}_t^{(0)}} - 1\right)\hat{z}_t^{(0)}. \tag{7.1.107}$$

第四步:计算

$$\hat{\boldsymbol{a}}(\hat{\boldsymbol{\alpha}}^{(0)}) = \left(\frac{\partial L(\boldsymbol{\theta})}{\partial \alpha_1},\cdots,\frac{\partial L(\boldsymbol{\theta})}{\partial \alpha_p}\right)^{\tau}_{\theta=\theta^{(0)}}, \tag{7.1.108}$$

$$\hat{\boldsymbol{a}}(\hat{\boldsymbol{\varphi}}^{(0)}) = \left(\frac{\partial L(\boldsymbol{\theta})}{\partial \varphi_0},\frac{\partial L(\boldsymbol{\theta})}{\partial \varphi_1},\cdots,\frac{\partial L(\boldsymbol{\theta})}{\partial \varphi_q}\right)^{\tau}_{\theta=\theta^{(0)}}. \tag{7.1.109}$$

根据(7.1.98)和(7.1.99)两式得第一步迭代公式:

$$\hat{\boldsymbol{\alpha}}^{(1)} = \hat{\boldsymbol{\alpha}}^{(0)} + [\hat{I}_{\alpha\alpha}^{(0)}]^{-1}\hat{\boldsymbol{a}}(\hat{\boldsymbol{\alpha}}^{(0)}), \tag{7.1.110}$$

$$\hat{\boldsymbol{\varphi}}^{(1)} = \hat{\boldsymbol{\varphi}}^{(0)} + [\hat{I}_{\varphi\varphi}^{(0)}]^{-1}\hat{\boldsymbol{a}}(\hat{\boldsymbol{\varphi}}^{(0)}). \tag{7.1.111}$$

第五步:如果迭代已进行到第 k 步,在(7.1.100)式、(7.1.103)式、(7.1.104)~(7.1.109)诸式中,用相应的第 k 步迭代所得的估计来代替,从而有第 $k+1$步迭代公式

$$\hat{\boldsymbol{\alpha}}^{(k+1)} = \hat{\boldsymbol{\alpha}}^{(k)} + [\hat{I}_{\alpha\alpha}^{(k)}]^{-1}\hat{\boldsymbol{a}}(\hat{\boldsymbol{\alpha}}^{(k)}), \tag{7.1.112}$$

$$\hat{\boldsymbol{\varphi}}^{(k+1)} = \hat{\boldsymbol{\varphi}}^{(k)} + [\hat{I}_{\varphi\varphi}^{(k)}]^{-1}\hat{\boldsymbol{a}}(\hat{\boldsymbol{\varphi}}^{(k)}). \tag{7.1.113}$$

迭代直到 $\|\hat{\boldsymbol{\alpha}}^{(k+1)}-\hat{\boldsymbol{\alpha}}^{(k)}\|$ 和 $\|\hat{\boldsymbol{\varphi}}^{(k+1)}-\hat{\boldsymbol{\varphi}}^{(k)}\|$ 达到预先指定的精度为止.

注意,在上述估计方法中,从理论上讲并不能保证每一步迭代都有 $\hat{\varphi}_j^{(k)}\geqslant 0\ (j=0,1,\cdots,q)$,并且 $1-\hat{\varphi}_1^{(k)}-\cdots-\hat{\varphi}_q^{(k)}>0$. 一般说来,为保证这些条件成立,应采用最优化数值方法来求带有约束的极大似然估计.

§7.2 非参数估计

在第3章中,我们对非线性时间序列模型进行分类时有两个出发点:一是按其预报函数的类型进行分类,即对平稳序列 $\{x_t\}$,其一步预报为

$$E\{x_t|x_{t-1},x_{t-2},\cdots\} = \varphi(x_{t-1},x_{t-2},\cdots),$$

并定义

$$e_t = x_t - E\{x_t | x_{t-1}, x_{t-2}, \cdots\},$$

则有

$$x_t = \varphi(x_{t-1}, x_{t-2}, \cdots) + e_t. \tag{7.2.1}$$

从统计分析角度看，由于实际中获得的观测样本总是有限的，因而很难对无穷多元预报函数 $\varphi(x_{t-1}, x_{t-2}, \cdots)$ 用有限多个样本去建模. 因此，实际中往往是考虑预报函数 φ 依赖于 $x_{t-1}, \cdots, x_{t-p}$ 等有限个历史值的情况，即

$$x_t = \varphi(x_{t-1}, \cdots, x_{t-p}) + e_t. \tag{7.2.2}$$

以(7.2.2)式作为(7.2.1)式的近似，这时 p 的取值应适当大. 当然，实际问题中的确存在预报函数 φ 仅依赖于有限个历史值的情况.

模型分类的另一个出发点是，从平稳 LARMA 模型直接推广而来，其一般形式为

$$x_t = \varphi(x_{t-1}, \cdots, x_{t-p}, \varepsilon_t, \cdots, \varepsilon_{t-q}) + \varepsilon_t. \tag{7.2.3}$$

§7.1中介绍的建模方法主要是针对非线性模型(7.2.2)和(7.2.3)，而且是函数 φ 仅被有限个参数所决定的情况——即参数型非线性模型. 这时，统计建模的任务主要是利用样本估计出模型中的未知参数. 本节要讨论模型(7.2.2)中 φ 为任意函数时如何利用样本去估计 φ，即所谓的非参数估计. 对于模型(7.2.3)，由于 φ 还是不可观测的噪声序列 $\{\varepsilon_t\}$ 的函数，因此其估计要比模型(7.2.2)更困难，在文献中尚未有成熟的估计方法. 以下仅讨论可加噪声 NLAR 模型的非参数估计问题，即

$$x_t = \varphi(x_{t-1}, \cdots, x_{t-p}) + \varepsilon_t, \tag{7.2.4}$$

式中 $\{\varepsilon_t\}$ 为白噪声序列，满足(3.2.2)式.

设 $x_1, \cdots, x_n$ 是模型(7.2.4)的一段样本. 利用这段样本估计函数 φ 有两种途径：一是利用计算机显示样本 $(x_1, \cdots, x_n)$ 的图形，例如样本数据对 (x_t, x_{t-1}) 在平面上的图形，(x_t, x_{t-1}, x_{t-2}) 在三维空间的图形等，根据样本 $x_1, \cdots, x_n$ 的有关图形的特征，寻找一个适当的函数 $\hat{\varphi}$ 作为 φ 的近似，这一方法难度大，不易实现. 况且这只是一种直观的探索性方法，故在这里不详加讨论. 另一途径是利用非参数回归中的非参数方法来估计(7.2.4)式中的函数 φ. 这是本节讨论的主要内容.

首先，我们考虑一阶可加噪声 NLAR 模型

$$x_t = \varphi(x_{t-1}) + \varepsilon_t, \tag{7.2.5}$$

这时，

$$\varphi(x_{t-1}) = E\{x_t | x_{t-1}\}. \tag{7.2.6}$$

为叙述方便，不妨假定 x_t 有密度函数 $f(x)$，记 $g(y,x)$ 为 (x_t,x_{t-1}) 的联合密度，$f(y|x)$ 为给定 $x_{t-1}=x$ 时，x_t 的条件密度函数，则由(7.2.6)式有

$$\begin{aligned}\varphi(x)&=E\{x_t|x_{t-1}=x\}=\int yf(y|x)\mathrm{d}y\\&=\int yg(y,x)\mathrm{d}y/f(x),\end{aligned}\tag{7.2.7}$$

式中 $f(x)$ 为 x_{t-1} 的边际密度函数，即

$$f(x)=\int g(y,x)\mathrm{d}y.$$

设 $x_1,\cdots,x_n$ 是模型(7.2.5)的一段样本，对每个 x，用 I_x 记以 x 为中心，长为 h_n 的区间，即 $I_x=\left[x-\frac{h_n}{2},x+\frac{h_n}{2}\right]$，则

$$P(x_{t-1}\in I_x)=\int_{x-h_n/2}^{x+h_n/2}f(z)\mathrm{d}z.$$

因此，$P(x_{t-1}\in I_x)$ 可用

$$\frac{1}{n}\sum_{t=1}^{n}I(x_{t-1}\in I_x)$$

估计. 从而 $P(x_{t-1}\in I_x)/h_n$ 可用

$$\frac{1}{nh_n}\sum_{t=1}^{n}I(x_{t-1}\in I_x)$$

估计. 当 h_n 充分小时，$P(x_{t-1}\in I_x)/h_n$ 可作为 $f(z)$ 在点 x 处的取值的估计. 故选择适当的正数 h_n，取

$$f_n(x)=\frac{1}{nh_n}\sum_{t=1}^{n}I(x_{t-1}\in I_x)\tag{7.2.8}$$

为 $f(x)$ 的估计.

类似地，$\int yg(y,x)\mathrm{d}y$ 可用

$$\frac{1}{nh_n}\sum_{t=1}^{n}I(x_{t-1}\in I_x)x_t\tag{7.2.9}$$

估计. 这样，由(7.2.7)～(7.2.9)式，φ 在点 x 处的取值 $\varphi(x)$ 可用

$$\varphi_n(x)=\frac{1}{nh_nf_n(x)}\sum_{t=1}^{n}I(x_{t-1}\in I_x)x_t\tag{7.2.10}$$

估计. $\varphi_n(x)$ 称为 $\varphi(x)$ 的*局部条件均值估计*. 令

$$W_{nt}(x)=\frac{I(x_{t-1}\in I_x)}{nh_nf_n(x)}\quad(t=1,2,\cdots,n),\tag{7.2.11}$$

则(7.2.10)式可写成

$$\varphi_{rn}(x) = \sum_{t=1}^{n} W_{nt}(x) x_t, \qquad (7.2.12)$$

且

$$W_{nt}(x) \geqslant 0 \quad (t = 1, \cdots, n), \ \sum_{t=1}^{n} W_{nt}(x) = 1. \qquad (7.2.13)$$

由(7.2.13)式可见，(7.2.11)式定义的$\{W_{nt}, t=1,\cdots,n\}$实际上是一个权函数. 若将样本 $x_1, \cdots, x_n$ 中恰好落入区间 I_x 的那些样本记为 $x_{k_1}, \cdots, x_{k_m}$，则由(7.2.11)式可知，$W_{nt}(x)$满足

$$W_{nt}(x) = \begin{cases} \dfrac{1}{m}, & \text{当 } x_{t-1} \in I_x; \\ 0, & \text{其他}. \end{cases} \qquad (7.2.14)$$

(7.2.14)式表明，用(7.2.12)式估计 φ 在点 x 处的取值时，只有样本$(x_{k_1+1}, x_{k_1}), \cdots, (x_{k_m+1}, x_{k_m})$起作用，且每对样本起的作用一样，而其余样本并不起作用. 直观上可以设想，为估计 φ 在点 x 处的取值 $\varphi(x)$，与 x 靠近的样本所起的作用应比远离 x 的样本大些，因此估计式(7.2.12)是合理的.

令

$$W(x) = I_{[-\frac{1}{2}, \frac{1}{2}]}(x) = \begin{cases} 1, & \text{当 } -\dfrac{1}{2} \leqslant x \leqslant \dfrac{1}{2}; \\ 0, & \text{其他}. \end{cases} \qquad (7.2.15)$$

由(7.2.11)式知

$$W_{nt}(x) = \frac{1}{nh_n f_n(x)}$$

$$= W\left(\frac{x - x_{t-1}}{h_n}\right) \quad (t = 1, \cdots, n), \qquad (7.2.16)$$

从而(7.2.12)式为

$$\varphi_{rn}(x) = \sum_{t=1}^{n} W\left(\frac{x - x_{t-1}}{h_n}\right) x_t. \qquad (7.2.17)$$

由(7.2.15)式可知，$W(x)$是 R^1 上的一个均匀密度函数. 如果将 $W(x)$取为更一般的密度函数，便得到熟知的核回归估计，其定义如下：

仍设 $x_1, \cdots, x_n$ 是来自模型(7.2.5)的样本. 设 $K(x)$是 R^1 上的一个给定的概率密度函数，h_n 是一个与 n 有关的常数，定义

$$\varphi_{kn}(x) = \sum_{t=1}^{n} K\left(\frac{x - x_{t-1}}{h_n}\right) x_t \bigg/ \sum_{t=1}^{n} K\left(\frac{x - x_{t-1}}{h_n}\right). \qquad (7.2.18)$$

$\varphi_{kn}(x)$称为$\varphi(x)$的核估计,h_n称为窗宽,$K(x)$称为核函数.令

$$W_{nt}(x)=K\left(\frac{x-x_{t-1}}{h_n}\right)\Big/\sum_{s=1}^{n}K\left(\frac{x-x_{s-1}}{h_n}\right) \tag{7.2.19}$$

$$(t=1,\cdots,n).$$

易见$W_{nt}(x)$满足(7.2.13)式,且

$$\varphi_{kn}(x)=\sum_{t=1}^{n}W_{nt}(x)x_t. \tag{7.2.20}$$

因此,不论是局部条件均值估计,还是核估计,都是一种权函数估计法.非参数回归的权函数估计法最初是由 Stone 提出的(见文献[97]).

从(7.2.18)式可知,核估计既同样本有关,又同核函数$K(x)$和窗宽h_n的选取有关.在给定样本后,一个核估计的统计性能的好坏,主要取决于核函数和窗宽的选取.至于如何选取h_n和$K(x)$,已有大量文献讨论了,这里不再详加讨论,有兴趣的读者可参阅文献[15]、[86]、[88]等.需要说明的是:对时间序列的非参数估计进行统计性质的研究,要借助于相依随机变量的极限理论.常用的相依性概念有两类:一类是α-混合相依,即,如果

$$\sup_{A\in\mathscr{F}_{1,n}^{x},B\in\mathscr{F}_{n+k,\infty}^{x}}|P(A\cap B)-P(A)P(B)|\leqslant\alpha_k,$$

而且当$k\to\infty$时,$\alpha_k\to0$,则称$\{x_t\}$是α-混合相依的.另一类是φ-混合相依,即

$$\sup_{\substack{A\in\mathscr{F}_{1,n}^{x},B\in\mathscr{F}_{n+k,\infty}^{x}\\P(A)>0}}|P(B|A)-P(B)|\leqslant\varphi_k,$$

而且当$k\to\infty$时,$\varphi_k\to0$,则称$\{x_t\}$是φ-混合相依的.两种相依性的关系表现为

$$\alpha_k\leqslant\varphi_k,\quad\forall\,k\geqslant1.$$

一般说来,要验证一个平稳序列$\{x_t\}$是α-混合相依的,还是φ-混合相依的,确实比较困难.但是,如果$\{x_t\}$是几何遍历的,则$\{x_t\}$一定是α-混合相依的,且α_k具有几何收敛速度(见文献[42]).

估计模型(7.2.5)中$\varphi(x)$的另一种方法是局部多项式回归方法,令

$$S(\boldsymbol{c})=\sum_{t=1}^{n}[x_t-\boldsymbol{c}^{\tau}\boldsymbol{U}_{tn}]^2K\left(\frac{x-x_{t-1}}{h_n}\right), \tag{7.2.21}$$

式中K是核函数,h_n是窗宽,$\boldsymbol{c}\in R^l$,且

$$\boldsymbol{U}_{tn}=\boldsymbol{F}(u_{tn}),$$

$$\boldsymbol{F}(u)=(1,u,\cdots,u^{l-1}/(l-1)!)^{\tau},$$

$$u_{tn} = (x - x_{t-1})/h_n.$$

关于 $\boldsymbol{c}$ 极小化 $S(\boldsymbol{c})$，并记其解为 $\boldsymbol{c}_n(x)$，即

$$S(\boldsymbol{c}_n(x)) = \inf_{c\in R^l} S(\boldsymbol{c}).$$

以下式作为 $\varphi(x)$ 的估计：

$$\varphi_{pn}(x) = \boldsymbol{c}_n(x)^\tau \boldsymbol{F}(0), \tag{7.2.22}$$

$\varphi_{pn}(x)$ 称为 $\varphi(x)$ 的*局部多项式回归估计*. Tsybakov(1986)及 Härdle 和 Tsybakov(1995)在 α-混合相依下，证明了 $\varphi_{pn}(x)$ 的渐近正态性.

到此为止，我们讨论了模型(7.2.5)的非参数自回归估计问题. 但是无论从理论上还是从应用上，高维数据的非参数拟合都具有重要意义. 下面我们讨论 NLAR(p)模型(7.2.4)中函数 φ 的非参数估计问题，其中假定 $\{e_t\}$ 为白噪声序列，即

$$x_t = \varphi(x_{t-1},\cdots,x_{t-p}) + \varepsilon_t, \tag{7.2.23}$$

式中 $\{\varepsilon_t\}$ 是白噪声序列，满足(3.2.2)式.

设 $x_1,\cdots,x_n$ 是模型(7.2.23)的一段样本. 条件均值函数

$$\varphi(x_{t-1},\cdots,x_{t-p}) = E\{x_t | x_{t-1},\cdots,x_{t-p}\}$$

的局部条件均值估计是前面叙述的(7.2.5)式中 $\varphi(x)$ 的条件均值估计的推广. 选择 $\delta_n>0$，且当 $n\to\infty$ 时，$\delta_n\to 0$. 对任意的 $\boldsymbol{x}=(x_1,\cdots,x_p)^\tau\in R^p$，令

$$\boldsymbol{x}_{t-1} = (x_{t-1},\cdots,x_{t-p})^\tau$$

和

$$I_n(\boldsymbol{x}) = \{i: 1\leqslant i\leqslant n, 且\ \|\boldsymbol{x}_{i-1} - \boldsymbol{x}\| < \delta_n\},$$

$$N_n(\boldsymbol{x}) = \# I_n(\boldsymbol{x}) = \{I_n(\boldsymbol{x})\ 中包含样本\ x_1,\cdots,x_n\ 的个数\}.$$

以

$$\hat{\varphi}_{cn}(\boldsymbol{x}) = \frac{1}{N_n(\boldsymbol{x})}\sum_{i\in I_n(\boldsymbol{x})} x_i \tag{7.2.24}$$

作为 $\varphi(\boldsymbol{x})$ 的局部条件均值估计.

Truong 和 Stone(1992)、Tran(1993)等讨论了在 α-混合相依下的条件均值估计 $\hat{\varphi}_{cn}(\boldsymbol{x})$ 在 L_2 范数和 L_∞ 范数下的最优收敛速度(见文献[109]和[111]). Truong 讨论了 $\hat{\varphi}_{cn}(\boldsymbol{x})$ 的强相容性和渐近正态性(见文献[110]).

对模型(7.2.23)中函数 φ 的核估计，也是前面同类估计的直接推广. 选定 R^p 上的核函数 $K_p(\boldsymbol{x})$ 及窗宽 h_n，满足 $n\to\infty$ 时，$h_n\to 0$. 对任意的 $\boldsymbol{x}\in R^p$，以

$$\hat{\varphi}_{kn}(\boldsymbol{x}) = \frac{\sum_{t=1}^{n} K_p\left(\frac{\boldsymbol{x} - \boldsymbol{x}_{t-1}}{h_n}\right) x_t}{\sum_{t=1}^{n} K_p\left(\frac{\boldsymbol{x} - \boldsymbol{x}_{t-1}}{h_n}\right)} \tag{7.2.25}$$

作为 $\varphi(\boldsymbol{x})$的核估计.

Auestad 和 Tjφstheim、Härdle 和 Vieu、Robinson 等(见文献[10]、[62]、[89])提出如下的乘积核估计:即选定R^1上的核函数 $K(x)$及 p 个窗宽参数 $h_{ni}(i=1,\cdots,p)$,满足 $n\to\infty$时,$h_{ni}\to 0\ (i=1,\cdots,p)$. 对任意的 $\boldsymbol{y}\in R^p$,记 $\boldsymbol{y}=(y_1,\cdots,y_p)^\tau$,令

$$\tilde{\varphi}_{kn}(\boldsymbol{y}) = \frac{\sum_{t=1}^{n}\prod_{i=1}^{p} K\left(\frac{y_i - x_{t-i}}{h_{ni}}\right) x_t}{\sum_{t=1}^{n}\prod_{i=1}^{p} K\left(\frac{y_i - x_{t-i}}{h_{ni}}\right)}. \tag{7.2.26}$$

以 $\tilde{\varphi}_{kn}(\boldsymbol{y})$作为 $\varphi(\boldsymbol{y})$的核估计.

关于核估计的大样本性质,Robinson、Härdle 和 Vieu 等证明了在 α-混合相依下,核回归估计的强相容性和渐近正态性(见文献[89]、[62]).

在考虑 NLAR(p)模型(7.2.23)的非参数估计时,随着模型阶数的增加,也即 $\boldsymbol{x}_t=(x_t,\cdots,x_{t-p+1})^\tau$ 维数的增加,将不可避免地遇到"维数灾难"的困扰,这一点在§6.3中已有过讨论. 因此,根据实际问题的需要,研究 NLAR(p)模型(7.2.23)的非参数估计的降维技术,是一个倍受重视的研究课题. 目前,文献中提出的 PP(Projection Pursuit)方法、逆回归方法和可加一元函数逼近方法,都是针对"维数灾难"提出的. 下面我们介绍模型(7.2.23)中 $\varphi(\boldsymbol{x})$的可加一元函数逼近方法. 其思想是:用一元非线性函数 $\varphi_i(x)$之和逼近多元函数 $\varphi(x_1,\cdots,x_p)$,即

$$\varphi(x_1,\cdots,x_p) \approx \mu + \sum_{i=1}^{p}\varphi_i(x_i). \tag{7.2.27}$$

故称之为*一元函数逼近方法*. 这样,可考虑第3章中的 NLAR 模型

$$x_t = \mu + \sum_{i=1}^{p}\varphi_i(x_{t-i}) + \varepsilon_t. \tag{7.2.28}$$

这时,统计建模的目的就是利用样本 $x_1,\cdots,x_n$ 估计上式中的 μ 和一元函数 $\varphi_1(x_1),\cdots,\varphi_p(x_p)$. 下面介绍由 Breiman 和 Friedman 提出的一种估计方法(见文献[17]). 这一方法称为*交替条件均值算法*(Alternating Conditional Expectation Algorithm),简称 ACE 算法.

为了叙述 ACE 算法,先由(7.2.28)式得到

$$\varphi_i(y) = E\{x_t - \sum_{j\neq i}\varphi_j(x_{t-j})\,|\,x_{t-i} = y\}, \qquad (7.2.29)$$

把式中的 $x_t - \sum_{j\neq i}\varphi_j(x_{t-j})$ 看作条件响应变量，再用非参数方法估计 φ_i. 在实际应用中，由于 $\varphi_i(i=1,\cdots,p)$ 都是未知的，因此，需要利用递推算法去估计 φ_i，直到所有估计都收敛为止. 其算法可表述如下：

(i) 令 $\varphi_{10}(y_1)=0,\cdots,\varphi_{p0}(y_p)=0$；

(ii) 给定精度 $\delta>0$ 和最大循环步长 M，对 $m=1$ 到 M 和 $k=1$ 到 p 用以下公式迭代计算：

$$\varphi_{km}(y_k) = E\left\{x_t - \sum_{j\neq k}\varphi_{j(m-1)}(x_{t-j})\,|\,x_{t-k} = y_k\right\}. \qquad (7.2.30)$$

如果

$$|\varphi_{km}(y_k) - \varphi_{k(m-1)}(y_k)| < \delta \quad (k = 1,\cdots,p), \qquad (7.2.31)$$

则停止迭代，结束 ACE 算法；否则，继续迭代，直到(7.2.31)式成立为止.

在实际问题中，用前面介绍的一阶 NLAR 模型的非参数估计来代替(7.2.30)中的条件期望，并记为 $\hat{\varphi}_{km}(y_k)$. 例如用核估计方法，则(7.2.30)式成为

$$\hat{\varphi}_{km}(y_k) = \frac{\sum_{t=1}^{n} K\left(\frac{y_k - x_{t-k}}{h_n}\right)\left[x_t - \sum_{j\neq k}\hat{\varphi}_{j(m-1)}(x_{t-j})\right]}{\sum_{t=1}^{n} K\left(\frac{y_k - x_{t-k}}{h_n}\right)}. \qquad (7.2.32)$$

到此为止. 我们已经看到，用 ACE 算法解决 ANLAR(p)模型(7.2.28)中的 φ_i 的非参数估计时，虽然其阶数 $p>1$，但是却只用到一维非参数拟合的算法，从而回避了"维数灾难"问题.

§7.3　模型阶数的估计

在时间序列的统计分析中，对参数型非线性模型的选择问题，包含有两项内容：一是确定模型的阶数，具体地说，就是确定条件期望 $E\{x_t|x_{t-1},\cdots,x_{t-p}\}$ 中的最大延迟 p. 二是确定参数型非线性模型的形式. 对非参数型的非线性模型而言，使用§7.2所介绍的非参数方法估计条件期望函数时，模型的选择只包括确定阶数的问题. 下面先讨论参数型非线性模型阶数的估计问题.

在线性时间序列分析中,最常用的,而且最有效的定阶方法是 AIC 准则方法.它是由 Akaike 首先提出(见文献[8]),其一般形式为

AIC=−2log(模型最大似然度)+2(模型的自由参数个数).

以最小化上述准则函数的解作为模型阶数的估计.这是一个很广泛的准则,它已被用于统计学的许多邻域.在线性时间序列分析中,主要用于确定模型阶数和模型的变元选择.例如对 LAR 模型,AIC 准则的形式为

$$\mathrm{AIC}(k)=\log\hat{\sigma}_k^2+2k/n, \tag{7.3.1}$$

其中 $\hat{\sigma}_k^2$ 是用某种估计方法对 ε_t 的残差方差的估计,而且是拟合 LAR(k) 的模型的拟合残差的方差.但是,最小化(7.3.1)式确定的阶数估计不具有相容性(见文献[95]),而且在理论上还表明:AIC 准则方法往往给出偏高的估计.因此文献中给出一些 AIC 准则的改进形式,例如

$$\mathrm{BIC}(k)=\log\hat{\sigma}_k^2+ck\log n/n, \tag{7.3.2}$$

$$\varphi(k)=\log\hat{\sigma}_k^2+ck\log\log n/n, \tag{7.3.3}$$

它们所确定的阶数都是真阶的相容估计(见文献[3]和[59]).以上定阶准则同样可以用于参数型非线性模型.下面通过门限自回归模型介绍其阶数的估计方法.

考虑如下门限自回归模型

$$x_t=\begin{cases}\varphi_{10}+\sum\limits_{j=1}^{p_1}\varphi_{1j}x_{t-j}+\varepsilon_{1t}, & \text{当 } x_{t-d}\leqslant r;\\ \varphi_{20}+\sum\limits_{j=1}^{p_2}\varphi_{2j}x_{t-j}+\varepsilon_{2t}, & \text{当 } x_{t-d}>r,\end{cases} \tag{7.3.4}$$

其中自回归系数 $\varphi_{1j}(j=0,1,\cdots,p_1)$ 和 $\varphi_{2j}(j=0,1,\cdots,p_2)$ 以及白噪声序列$\{\varepsilon_t\}$满足例7.1.6的假定.

设 $x_1,\cdots,x_n$ 是模型(7.3.4)的一段样本,利用例7.1.6所述的估计方法,在给定 p_1 和 p_2 下,分别估计出模型(7.3.4)的两个子模型的残差方差:

$$\hat{\sigma}_{jn}^2=\frac{Q_{jn}(\hat{\theta}_n)}{n_j}\quad(j=1,2), \tag{7.3.5}$$

式中 Q_{jn}、$\hat{\theta}_n$ 和 n_j 的定义见例7.1.6.这时相应于每个子模型的 AIC 统计量分别为

$$\mathrm{AIC}(p_1)=\log\hat{\sigma}_{1n}^2+2(p_1+1)/n_1,$$

$$\mathrm{AIC}(p_2)=\log\hat{\sigma}_{2n}^2+2(p_2+1)/n_1.$$

对于预先给定的模型阶数的上界 L,分别关于 p_1 和 p_2 极小化 AIC(p_1)和

AIC(p_2)，即求 $\hat{p}_1$和 $\hat{p}_2$，使得

$$\text{AIC}(\hat{p}_1) = \min_{1 \leqslant p_1 \leqslant L} \text{AIC}(p_1),$$

$$\text{AIC}(\hat{p}_2) = \min_{1 \leqslant p_2 \leqslant L} \text{AIC}(p_2).$$

由此得模型(7.3.4)的阶的估计 $\hat{p}_1$和 $\hat{p}_2$.

将上述 AIC(p_1)和 AIC(p_2)换成 BIC(p_1)和 BIC(p_2)，同样可得到模型(7.3.4)的阶的估计.

现在讨论非参数型非线性时间序列模型的定阶问题. 按一步预报性质，非线性时间序列模型可表为(7.2.1)式的形式，即

$$x_t = \varphi(x_{t-1}, x_{t-2}, \cdots) + e_t, \tag{7.3.6}$$

其中 e_t 和 φ 与(7.2.1)式含义相同. 当 φ 为非参数型函数时，模型(7.3.6)为非参数时间序列模型. 在实际数据分析中，常假定预报函数仅依赖于有限个历史值，如(7.2.2)式的形式，即

$$x_t = \varphi(x_{t-1}, \cdots, x_{t-p}) + e_t, \tag{7.3.7}$$

其中

$$\begin{aligned}\varphi(x_{t-1}, \cdots, x_{t-p}) &= E\{x_t | x_{t-1}, \cdots, x_{t-p}\} \\ &= E\{x_t | x_{t-1}, x_{t-2}, \cdots\}.\end{aligned}$$

由于实际建模时，并不知道模型(7.3.7)的阶数. 这时，自然要问，应根据什么原则来估计阶数 p 呢？一种较直观的想法是：在使得一步预报误差方差达到最小值的 p 中选择最小者. 具体地说，令 p 为模型(7.3.7)的真阶，它是未知的. 再令

$$e_t(k) = x_t - E\{x_t | x_{t-1}, \cdots, x_{t-k}\}, \tag{7.3.8}$$

$$\begin{aligned}\sigma^2(k) &= E\{e_t(k)\}^2 \\ &= E\{x_t - E\{x_t | x_{t-1}, \cdots, x_{t-k}\}\}^2.\end{aligned} \tag{7.3.9}$$

显然，对 $k \geqslant 1$ 有 $\sigma^2(k) \leqslant \sigma^2(k-1)$，其中 $\sigma^2(0) = Ex_t^2$. 按上述直观的想法，当 $k < p$ 时，$\sigma^2(k) > \sigma^2(p)$，所以不应选 $k < p$；当 $k > p$ 时，$\sigma^2(p) = \sigma^2(k)$，所以也不应当选 $k > p$. 最后，只能选 $k = p$.

Cheng 和 Tong(见文献[35])对模型(7.3.7)，提出了一种相容的定阶方法，称为 Cross-Validation 方法.

对模型(7.3.7)的一段样本 $x_1, \cdots, x_n$，令 $\boldsymbol{x}_t(k) = (x_{t-1}, \cdots, x_{t-k})^{\tau}$，定义

$$\text{CV}(k) = \frac{1}{n-k} \sum_{t=k+1}^{n} \{x_t - \hat{\varphi}_{-t}(\boldsymbol{x}_t(k))\}^2 W(\boldsymbol{x}_t(k)), \tag{7.3.10}$$

其中 $\hat{\varphi}_{-t}$ 是去掉第 t 个数据 x_t 后，对自回归函数 φ 的核估计，注意其阶数 p 用尝试阶数 k 代替，其估计为

$$\hat{\varphi}_{-t}(y)=\frac{\sum\limits_{\substack{s=k+1\\ s\neq t}}^{n}\prod\limits_{i=1}^{k}K\left(\frac{y_i-x_{s-i}}{h_n}\right)x_s}{\sum\limits_{\substack{s=k+1\\ s\neq t}}^{n}\prod\limits_{i=1}^{k}K\left(\frac{y_i-x_{s-i}}{h_n}\right)}.$$

(7.3.10)式中 W 为适当选取的非负权函数. 对预先给定的模型(7.3.7)阶的上界 L，在 $\{1,2,\cdots,L\}$ 上极小化 $\mathrm{CV}(k)$，这时模型阶的估计 $\hat{p}$ 满足

$$\mathrm{CV}(\hat{p})=\min_{1\leqslant k\leqslant L}\mathrm{CV}(k). \qquad (7.3.11)$$

Cheng 和 Tong 证明了在适当的条件下，由(7.3.11)式求得的 $\hat{p}$ 是真阶 p 的相容估计.

在本节结束前，简单讨论一下关于参数型非线性时间序列模型的类型判别问题. 由于非线性时间序列模型的种类繁多，如见第3章所述，解决类型判别问题很难. 即使限定适当的类型范围，例如只考虑可加噪声 NLAR 模型，这仍是很困难的问题. 因为一般可加噪声 NLAR(p)模型的形式为

$$x_t=\varphi(x_{t-1},\cdots,x_{t-p};\boldsymbol{\theta})+\varepsilon_t, \qquad (7.3.12)$$

其中 φ 是仅依赖有限参数向量 $\boldsymbol{\theta}$ 的参数型自回归函数. 可想而知，其可能的类型有多么广泛. 在线性情况下，当 p 已知时，φ 的函数类型就完全确定了，没有什么选择问题. 而在非线性情况下，不仅存在选择类型的问题，而且是实际应用者非常关心的问题. 然而，目前尚无有效方法解决这一问题. 在此，我们推荐几种现实的方法如下：

一种是从实际背景中确定函数类型，例如指数型自回归模型来自振动的物理背景.

另一种方法是：用图形方法观察数据 $x_1,\cdots,x_n$ 的低维结构. 如在§7.2中提到的，观察(x_t,x_{t-1}) $(t=1,2,\cdots,n)$在二维平面上的趋势，以猜测 φ 的可能类型. 这种方法只能用于低阶情况，但它能发挥研究者的观察判断能力.

最后一种是：当我们对 φ 的类型有初步认识或猜测时，例如有了一个或两个重点选择的模型时，可用第6章中的假设检验方法对它们逐一检验，选出能接受的模型类型.

第 8 章

非线性时间序列与相关领域的联系

非线性时间序列分析与近代某些热点研究领域有密切联系，联系的纽带就是非线性性. 从广义上说，它们都是属于非线性科学的分支领域. 本章要简述非线性时间序列与混沌、分形几何和神经网络等领域的联系.

§8.1 非线性时间序列与混沌

在许多学科领域，混沌已成为被非常关注的研究方向，比如在物理、力学、生物、经济等领域，此外还包括数学和统计学领域. 在不同的领域中，混沌的含义和研究方法有所不同，以至对混沌的概念尚未有统一的定义. 尽管如此，混沌的起源是与“非线性”分不开的.

在数学领域中，非线性微分方程和非线性差分方程都是描述非线性动态系统的重要分支学科，因此，它们必然涉及混沌概念. 而非线性时间序列与非线性差分方程有着直接的密切联系，所以也必然与混沌有联系. 在这一节所叙述的混沌与时间序列的联系，只涉及离散动态系统中的混沌概念. 为了突出介绍混沌概念与时间序列的联系，本节仅讨论一维动态系统中的混沌概念.

一、一维非线性动态系统的复杂性

在介绍一维非线性动态系统的复杂性时，我们限于取值在[0,1]区间上的系统. 将要介绍的混沌概念，就是起源于非线性系统的复杂性. 在第 2 章中，我们已经看到了非线性系统比线性系统复杂得多. 在那里引进了吸引元、吸引域和吸引元的稳定性等概念，以刻划非线性系统的复杂结构. 还通过若干非线性系统的例子展现出这些概念所具有的丰富多彩的内容. 在此我们将从另一角度描述非线性系统的复杂性，同样表现出富有诱惑力的结构特征. 这些内容与即将介绍的混沌概念有着更直接的联系.

现在讨论从[0,1]到[0,1]的变换 φ，仍使用前几章的记号表示由 φ

所决定的动态系统，即

$$\begin{cases} x_t = \varphi(x_{t-1}) \quad (t \geqslant 1); \\ x_0 \in [0,1]. \end{cases} \tag{8.1.1}$$

并以 $\varphi_n(x)$ 表示从 $x_0=x$ 出发，依上式迭代 n 步的取值.

对于某些函数 φ 而言，当 x 和 y 适当靠近时，随着 n 的不断增大，$\varphi_n(x)$ 和 $\varphi_n(y)$ 可以任意接近. 确切地说，对于任意给定的 $\delta>0$，存在 $\varepsilon>0$ 和 n_0，使得当 $|x-y|\leqslant\varepsilon$ 时，必有

$$|\varphi_n(x) - \varphi_n(y)| \leqslant \delta, \quad 当 n \geqslant n_0. \tag{8.1.2}$$

例如，当 $\varphi(x)=\alpha x$，且 $|\alpha|<1$ 时，则 $\varphi_n(x)=\alpha^n x$，从而

$$|\varphi_n(x) - \varphi_n(y)| \leqslant |\alpha|^n \cdot |x-y|.$$

由此可见，对于给定的 $\delta>0$，取 $\varepsilon=\delta, n_0=1$，则由上式立刻得知(8.1.2)式成立.

但是，许多非线性系统不具有上述的性质. 换言之，无论两个不同的初始值 x 和 y 如何接近，从它们出发的迭代序列 $\varphi_n(x)$ 和 $\varphi_n(y)$，随着 n 的增大并不能任意靠近. 这一特征被泛称为系统(8.1.1)对初值是灵敏的，或简称为*初值灵敏*，其严格定义如下：

定义 8.1.1 系统(8.1.1)被称为*初值灵敏*的，如果存在 $\delta>0$ 使得对任意 $x\in[0,1]$ 以及 x 的任意一个领域 $V(x)$，总存在 $y\in V(x)$ 和正整数 n，使得 $|\varphi_n(x)-\varphi_n(y)|>\delta$. 此时又称 φ 是*初值灵敏*的.

例 8.1.2 考虑(2.2.13)式所确定的系统(8.1.1)，即

$$\begin{cases} x_t = 2x_{t-1}\bmod(1) \quad (t \geqslant 1); \\ x_0 \in [0,1]. \end{cases} \tag{8.1.3}$$

为了说明此系统中的 $\varphi(x)=2x\bmod(1)$ 是初值灵敏的，取 $\delta=\frac{1}{2}$，并先考虑 $x=\frac{1}{2}$ 以及 $\frac{1}{2}$ 的领域

$$V\left(\frac{1}{2}\right) = \left\{z: \left|z-\frac{1}{2}\right| \leqslant \left(\frac{1}{2}\right)^{100}\right\},$$

易见，取 $y=\frac{1}{2}+\frac{\sqrt{2}}{(10)^{101}}\in V\left(\frac{1}{2}\right)$，则由(8.1.3)式知，$\varphi_n\left(\frac{1}{2}\right)=0$ 对每个 $n\geqslant1$ 成立；而 y 是无理数，从而不难看出，有无穷多个 $\varphi_n(y)$ 大于 $\frac{1}{2}$. 于是有无穷多个 n 满足下式：

$$\left|\varphi_n\left(\frac{1}{2}\right) - \varphi_n(y)\right| = |\varphi_n(y)| \geqslant \frac{1}{2} = \delta.$$

这就表明,(8.1.2)式不能成立.

对于 $x=\frac{1}{2}$ 的其他领域,依类似方法仍可证明(8.1.2)式不能成立,只要选取 y 为足够靠近 $\frac{1}{2}$ 的无理数即可证明此结论.

对于 $x\in[0,1]$,在 $x\neq\frac{1}{2}$ 时,仍可仿照上述推理证明(8.1.2)式不能成立.这里不再赘述.

例 8.1.3　考虑(8.1.1)中的 $\varphi(x)=\sin\pi x/2, x\in[0,1]$. 易见,$\varphi(0)=0,\varphi(1)=1$,而且对任意 $x_0\in(0,1)$,从 x_0 出发由 φ 迭代 n 次的值 $\varphi_n(x_0)$ 满足

$$\lim_{n\to\infty}\varphi_n(x_0)=1.$$

由此可知,此例中的 φ 不具有初值灵敏性,或者说,φ 不是初值灵敏的.

例 8.1.4　考虑(8.1.1)式中的 φ 为如下函数:

$$\varphi(x)=\begin{cases}2x, & \text{当 } 0\leqslant x\leqslant\frac{1}{4};\\ 2x-\frac{1}{2}, & \text{当 } \frac{1}{4}<x\leqslant\frac{3}{4};\\ 2x-1, & \text{当 } \frac{3}{4}<x\leqslant 1.\end{cases}\tag{8.1.4}$$

不难看出,上式中的 φ 是初值灵敏的,论证的方法与前面例 8.1.2 相似. 但是,例 8.1.2 和例 8.1.4 中的 φ 有一种本质性的差别. 为说明这种差别先要引进如下记号,记 $A\subset[0,1]$ 为一子集,并记 A 的象集为

$$\varphi(A)=\{y: y=\varphi(x), x\in A\}.$$

易见,由(8.1.4)式定义的 φ 满足

$$\varphi\left(\left[0,\frac{1}{2}\right]\right)=\left[0,\frac{1}{2}\right],\ \varphi\left(\left[\frac{1}{2},1\right]\right)=\left[\frac{1}{2},1\right].$$

现在可以指出(8.1.3)式和(8.1.4)式中的 φ 有何区别了. 对于(8.1.3)式中的 φ 而言,它不满足上式,反而满足

$$\varphi\left(\left[0,\frac{1}{2}\right]\right)=[0,1],$$

而(8.1.4)式中的 φ 却不满足上式. 实际上,系统(8.1.4)可以分别限制在 $\left[0,\frac{1}{2}\right]$ 和 $\left[\frac{1}{2},1\right]$ 两个子区间上而形成两个独立系统. 但是系统(8.1.3)不能作如此分解.

为了描写上述非线性系统之间的差异,需要给出如下的定义:

定义 8.1.5 系统(8.1.1)被称为瞬变的,如果对[0,1]中的任意两个开集 U 和 V,存在正整数 n 使得 $\varphi_n(U)\cap\varphi_n(V)$是非空的.此时也称 φ 为瞬变的.

粗略说,系统(8.1.1)为瞬变的,就意味着系统的轨道$\{\varphi_n(x),n\geqslant 1\}$将走遍[0,1]中的各个角落,至少对几乎所有的(在 Lebesgue 测度意义下)$x_0=x\in[0,1]$是这样的.不难验证,系统为瞬变的充分必要条件是,存在 $x\in[0,1]$,使得 $x_0=x$ 及$\{\varphi_n(x),n\geqslant 1\}$为[0,1]中的稠密子集.

按定义 8.1.5 重新考查例 8.1.2 和例 8.1.4 可知,(8.1.3)式中的 φ 是瞬变的,而(8.1.4)式中的 φ 不是瞬变的.由此可见,瞬变的概念描述了以上两个 φ 的某种本质的差别.

非线性系统的复杂性,还表现在它们的极限环方面,当系统中的 φ 为连续函数时,其复杂性还可表现在其周期解方面.为了说明这一点,我们将讨论以下的一个特殊的例子.

例 8.1.6 考虑如下的变换,又称 Logist 变换:

$$\varphi(x)=ax(1-x),\ x\in[0,1], \tag{8.1.5}$$

其中参数 a 满足 $0\leqslant a\leqslant 4$.显然,φ 是从[0,1]到[0,1]上的连续变换.根据第 2 章所述,由(8.1.5)式的 φ 所决定的系统(8.1.1)的极限点和极限环同 φ 的不动点和周期解分别重合.我们仍把不动点看作周期为 1 的周期解,下面讨论(8.1.5)式中 φ 的周期解与系数 a 的关系.

(i) 当 $0\leqslant a\leqslant 1$ 时,易见 $\varphi(x)=x$ 有唯一解,即 $\varphi(0)=0$.此解是 φ 的周期为 1 的周期解,按照第 2 章中关于吸引域的定义,易知不动点$\{0\}$的吸引域为[0,1],从而 φ 不再有其他周期解(参见性质 2.2.15).

(ii) 当 $1<a\leqslant 3$ 时,$\varphi(x)=ax(1-x)=x$ 有两个解,即 $x=0$ 或 $x=(a-1)/a$.不难看出,不动点$\{0\}$只吸引$\{0\}$和$\{1\}$两个点,而$\{(a-1)/a\}$的吸引域为(0,1).由此可见,此时 φ 没有其他吸引元.换言之,此时的 φ 仍然只有周期为 1 的周期解,与情况(i)相比,其个数多了 1 个,即$\{0\}$和$\{(a-1)/a\}$都是 φ 的周期为 1 的周期解.

(iii) 当 $3<a<1+\sqrt{6}$ 时,$x=0$ 和$(a-1)/a$ 仍是 φ 的两个不动点,但是$\{0\}$只吸引$\{0\}$和$\{1\}$两个点,$\{(a-1)/a\}$只吸引其自身.所以 φ 必有其他吸引元.考虑 $\varphi_2(x)=\varphi(\varphi(x))$,并求其不动点之解,即求解

$$\varphi_2(x)=a\{ax(1-x)\}\{1-ax(1-x)\}=x. \tag{8.1.6}$$

上式为一元四次方程,可写为

$$a_0x^4+a_1x^3+a_2x^2+a_3x+a_4=0. \tag{8.1.7}$$

显然,φ 的不动点也必是 φ_2 的不动点,所以有

$$\varphi_2(0)=\varphi(\varphi(0))=0;\ \varphi_2\left(\frac{a-1}{a}\right)=\varphi\left(\varphi\left(\frac{a-1}{a}\right)\right)=\frac{a-1}{a}. \tag{8.1.8}$$

此式表明:$\lambda_1=0,\lambda_2=(a-1)/a$ 是(8.1.7)的两个根.于是综合以上三式可得如下一元二次方程

$$a^3x^2+(\lambda_2-2)a^3x+a^3(1-\lambda_2)^2+a^2=0.$$

由此式得到(8.1.6)式的另外两个根为

$$\lambda_3=\{a+1+(a^2-2a-3)^{\frac{1}{2}}\}/2a,$$

$$\lambda_4=\{a+1-(a^2-2a-3)^{\frac{1}{2}}\}/2a.$$

在 $3<a<1+\sqrt{6}$ 时,易见 $\lambda_3\neq\lambda_4$,而且 $\varphi(\lambda_3)=\lambda_4,\varphi(\lambda_4)=\lambda_3$.这些事实表明 $\{\lambda_3,\lambda_4\}$ 是 φ 的周期为 2 的周期解.再利用推论 2.3.20 不难验证,此周期解为稳定的,而且其吸引域为 $[0,1]\backslash\{0,(a-1)/a,1\}$,即除去 $\{0\}$、$\{(a-1)/a\}$ 和 $\{1\}$ 以外的其余 $[0,1]$ 中各点都被吸引.

(iv) 当 $1+\sqrt{6}\leqslant a\leqslant 4$ 时,$x=0$ 和 $(a-1)/a$ 仍是 φ 的两个不动点.但是关于 φ 的周期解的讨论要复杂多了.下面要用到一个有趣的结果,为此先介绍自然数 $\{1,2,\cdots\}$ 的 Sarkovskii 序[92],即

$3,\ 5,\ 7,\cdots$

$2\cdot3,\ 2\cdot5,\ 2\cdot7,\cdots$

$2^2\cdot3,\ 2^2\cdot5,\ 2^2\cdot7,\cdots$

$2^3\cdot3,\ 2^3\cdot5,\ 2^3\cdot7,\cdots$

……………………………

……………………………$2^5,2^4,2^3,2^2,2,1.$

任给一个自然数,例如 28,此数可唯一地分解为 $28=2^2\cdot7$,于是自然数 28 在以上的 Sarkovskii 序中找到其位置,即在第三行左起第三个位置,是 $2^2\cdot7=28$.换言之,在上述排序中的第一行,即全体奇数排序,以及第二行——全体 2 倍的奇数,和第三行中的前两个数 $2^2\cdot3$ 和 $2^2\cdot5$,它们都在自然数 28 的前面,其余自然数都在 28 的后面.又如自然数 $16=2^4$,它在 Sarkovskii 序的倒数第 5 位.以上的排序被称为 Sarkovskii 序,因为他证明了:从 $[0,1]$ 到 $[0,1]$ 上的连续变换 φ,如果具有周期长度为 m 的周期解,则按照 Sarkovskii 序排在 m 之后的任何自然数 n,φ 必有长度为 n 的周期解.

将 Sarkovskii 序用到例 8.1.6 中来,并继续我们的讨论,可以发现随

着(8.1.5)式中的 a 值不断增大,φ 的周期解的个数,特别是不同长度的周期解的个数将单调增加.例如在 $0\leqslant a\leqslant 3$ 时,φ 只有周期长度为 1 的周期解.在 $3<a<1+\sqrt{6}$ 时,φ 有周期长度为 1 和 2 的周期解.当 a 再增大时,如增大至某个 a_3 时,则 φ 有长度为 1、2 和 $2^2=4$ 的周期解.依此类推,直到 a 增大到 $a_\infty=3.56994\cdots$ 时,φ 拥有长度为 $1,2,2^2,2^3,\cdots$ 的各种长度的周期解.当 a 超过 a_∞ 而继续增大时,φ 的周期解仍会增加.直至 a 增大到 $a'=3.839\cdots$ 时,φ 具有长度为 3 的周期解,按照 Sarkovskii 的前述结果,φ 具有长度为任何自然数的周期解.

讨论至此,关于 φ 的周期解仍有许多问题.例如,当参数 a 变大时,具有相同长度周期的周期解的个数如何变化的问题.又如,对于指定的 a,φ 的周期解中哪些是稳定的?其吸引域是什么?这些问题有的容易解决,有的仍待解决.这里不逐一讨论了,我们仅指出:当 $a=4$ 时,φ 的值域才是 $[0,1]$,而且 φ 的任何周期解都不是稳定的,事实上,$[0,1]$ 是 φ 的最大吸引元,而且其吸引域也是 $[0,1]$.下文将指出,此时的 φ 是混沌变换.

在这一小节最后顺便指出,上面提到的 Sarkovskii 的结果只适用于所有连续变换 φ,而不限于(8.1.5)式中的 φ 变换,但是对于不连续变换并不适用.例如,

$$\varphi(x)=\begin{cases}\dfrac{1}{3}, & \text{当 } x=0;\\ \dfrac{2}{3}, & \text{当 } x=\dfrac{1}{3};\\ 0, & \text{当 } x=\dfrac{2}{3};\\ x, & \text{其他 } x\in[0,1].\end{cases}$$

易见,$\left\{0,\dfrac{1}{3},\dfrac{2}{3}\right\}$ 是 φ 的周期解,其周期长度为 3,但是 φ 没有其他周期解.

二、混沌与随机性

如前所述,吸引集、吸引域和稳定性等概念能够描述非线性动态系统的某些结构的特征.但是,当吸引集只包括有限个点时,显然,根据前一小节的定义 8.1.1 和定义 8.1.5,这样的系统不具有初值灵敏性和瞬变性.由此可见,前一小节的初值灵敏性和瞬变性,是刻画非线性系统结构特征的另一类重要的概念.而且我们将看到,这些概念与混沌有密切的联系.

说起混沌概念,它起源于人们对某些非线性动态系统的惊奇发现,而

不是研究非线性系统理论时必须引入的抽象概念. 例如,在用非线性微分方程描述流体现象时,一个非常简单的确定性动力系统,却可以产生无法预测的行为,就像随机过程所呈现出的特征. 这种惊奇的发现,引起了许多物理、生物、经济以及统计方面的学者们的关注(如见文献[31]和[41]). 这种确定性系统,所表现出随机性态的现象不仅在流体力学研究中出现,在其他领域同样存在. 然而,在此类现象中的"随机性态"并不是有明确含义的术语,所以在用混沌概念描述这类现象时,出现了不同含义的混沌定义. 以下将简单介绍几种不同的混沌定义.

定义 8.1.7　设 φ 为从[0,1]到[0,1]上的变换,φ 称为混沌的,若:

(i) φ 是初值灵敏的;

(ii) φ 是瞬变的;

(iii) φ 的周期解的点集全体在[0,1]中稠密.

满足上述定义的函数 φ,可称为混沌变换. 由它决定的动态系统(8.1.1),表现出不可预测性,不可分解性,但又有一定的规律性,这些都与随机过程的性态相似. 以上定义对于连续和非连续函数都适用. 实际上,上述定义还可推广到更广的情况,例如 φ 为有界集 D 到 D 上的变换,又如 φ 为多元变换函数等,这里不作全面讨论了. 下面给出另一种混沌的定义.

定义 8.1.8　设 φ 为从[0,1]到[0,1]上的变换,如果 φ 是连续的,而且有长度为 3 的周期解,则称 φ 为混沌的.

为了说明此定义与"随机性态"有联系,先引入如下的符号序列的定义. 仍以 $\varphi_n(x)$ 表示从 $x_0=x\in[0,1]$ 出发,由 φ 迭代 n 步的取值,以 y_n 表示与 $\varphi_n(x)$ 相应的符号序列,即

$$y_n = 0,\text{当 } \varphi_n(x) < c;\quad y_n = 1,\text{当 } \varphi_n(x) \geqslant c.$$

这里的常数 c 是(0,1)中的某个值,例如 $c=\frac{1}{2}$. 在例 8.1.6 中,如果 φ 具有长度为 3 的周期解,即(8.1.5)式中的参数 $a>3.839\cdots$ 时,则对任意长度为无穷的 0－1 序列,均可找到[0,1]中的适当的初值 $x_0=x$,使得由 $\varphi_n(x)$ 决定的符号序列 $\{y_n\}$ 与指定的 0－1 值序列相重合. 此事实表明,由变换 φ 产生的迭代序列所决定的符号序列全体与 Bernoulli 试验所产生的0－1值序列相重合. 这便是定义 8.1.8 所表现出的"随机性态". 显然,它比定义 8.1.7 所表现的"随机性态"简单且间接.

我们还将用概率方法给出第三种混沌的定义. 为此,先介绍动态系统

的不变分布和遍历性概念. 现在将系统(8.1.1)看作如下的随机变量的迭代系统,即

$$\begin{cases} x_t = \varphi(x_{t-1}) \quad (t \geqslant 1); \\ x_0 \sim F_0, \end{cases} \tag{8.1.9}$$

式中 F_0 为[0,1]上的某个概率分布,初值 x_0 为服从 F_0 分布的随机变量. 于是由(8.1.9)式迭代产生的序列$\{x_t, t \geqslant 0\}$是一串随机变量序列,它们的分布记为 $F_0, F_1, \cdots$,显然 x_t 的分布 F_t 是被 φ 和 F_0 唯一决定的.

定义 8.1.9　概率分布 F 被称为(8.1.9)式中变换 φ 的不变分布,如果当 x_0 服从 F 分布时,$x_1=\varphi(x_0)$也服从分布 F. 从而所有 x_t 都服从分布 F,所以 F 也被称为系统(8.1.9)的**不变分布**.

定义 8.1.10　区间[0,1]中的子集 A 被称为变换 φ 的**不变子集**,如果 A 的象集 $\varphi(A)=A$.

定义 8.1.11　设 F 为 φ 的不变分布,称 F 为**遍历的**,如果对任何φ的不变集 A,必有 $P_F(A)=0$ 或 1,这里的 P_F 为由概率分布 F 在[0,1]上决定的概率测度. 换言之,若 $x_0 \sim F$,则 $P_F(A)=P(x_0 \in A)$.

定义 8.1.12　称系统(8.1.1)中的 φ 是混沌的,如果

(i) φ 是瞬变的;

(ii) φ 有一个遍历的分布 F,它与[0,1]上的 Lebesgue 测度是相互绝对连续的.

以上定义的混沌概念与随机概念的联系是直接的,具有遍历分布 F 的变换 φ 有如下的性质:

性质 8.1.13　如果 φ 有不变分布 F,而且 F 是遍历的,则对于任何关于分布 F 为可测的函数 g,只要 g 关于分布 F 是绝对矩有穷的,必有

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} g(\varphi_t(x)) = \int g(x) \mathrm{d}F(x), \quad \text{a.s. } x_0. \tag{8.1.10}$$

其中"a.s. x_0"是指对几乎所有的(依 Lebesgue 测度)$x_0 \in [0,1]$而言的.

现在讨论两个例子:首先考虑例 8.1.2 中的 $\varphi(x)=2x\bmod(1)$. 经过简单而直接的验证可知,[0,1]上的均匀分布是此变换的不变分布. 如果 A 是 φ 的不变子集,不难验证,A 的 Lebesgue 测度 $\mu(A)=0$ 或 1,从而可知[0,1]上的均匀分布是遍历的分布. 又因[0,1]上的均匀分布与[0,1]上的 Lebesgue 分布是相同的,从而必相互绝对连续. 综上所述可知,φ 是在定义 8.1.12 意义下为混沌的. 另外,前面曾指出,例 8.1.2 中的 φ 是初值灵敏的. 不难看出,φ 的周期点在[0,1]中是稠密的,所以 φ 也是按定义

8.1.7为混沌的. 但是由于 φ 不是连续函数，所以谈不上按照定义 8.1.8 所定义的混沌问题.

再考虑例 8.1.6 中的 $\varphi=4x(1-x)$，即参数 $a=4$ 的情况. 使用直接验证不难看出，Bata(0.5,0.5)分布为 φ 的一个不变分布，它与[0,1]上的 Lebesgue 分布也是相互绝对连续的. 同样可以验证此分布是 φ 的遍历的不变分布. 从而 φ 在定义 8.1.12 意义下是混沌的. 但是当 $3.839\cdots<a<4$ 时，$\varphi(x)=ax(1-x)$在定义 8.1.12 意义不是混沌的，因为 φ 的值域是[0,1]的真子集，从而不满足定义 8.1.12 中的第(ii)条. 同理，它也不是定义 8.1.7 意义下的混沌，不过它是在定义 8.1.8 意义下的混沌变换.

这里还须指出，以上定义的不变集合、不变分布与第 2 章的吸引元有密切联系. 为说明此事，不妨仅考虑从[0,1]到[0,1]上的连续变换的情况. 此时变换 φ 的不动点和周期解都是 φ 的不变子集. 每个周期解都对应着 φ 的一个不变分布. 例如 $\{c_1,c_2,\cdots,c_p\}$ 是 φ 的一个周期解集合，即

$$\varphi(c_i)=c_{i+1}\quad(i=1,2,\cdots,p-1);\ \varphi(c_p)=c_1.$$

则此集合是 φ 的不变子集，而且在 $c_1,c_2,\cdots,c_p$ 各点取相等概率——$1/p$ 的离散分布，是 φ 的一个不变分布. 由此可见，φ 的不变分布并不一定是唯一的. 不仅如此，遍历的不变分布也不一定唯一. 但是，两个遍历的不变分布或者是正交的，或是相互绝对连续的. 从而能与 Lebesgue 测度相互绝对连续的遍历的不变分布是唯一的，当然，这样的分布不是对任何 φ 都存在的. 如果 $F_1,F_2,\cdots$是 φ 的一串遍历的不变分布，显然

$$F=\sum_{j=1}^{\infty}\alpha_jF_j,\quad \sum_{j=0}^{\infty}\alpha_j=1\quad(\alpha_j\geqslant 0)\tag{8.1.11}$$

也是 φ 的不变分布，但不一定是遍历的，除非在(8.1.11)式中有某个 $\alpha_i=1$，其他均取零值.

三、时间序列与混沌

如前所述，混沌概念旨在刻画具有随机性态的确定性动态系统. 从概率论观点看，这里所称的“随机性态”缺少明确的含义. 在系统(8.1.1)的初值给定后，$x_1=\varphi(x_0)$并没有随机性. 即使迭代 n 步以后，$\varphi_n(x_0)$也是被 x_0 和 φ 唯一决定的，仍然没有随机性. 确切地说，前面所言的随机性态只能说是在某种意义下与随机过程相像. 另一方面，从现实角度看，所有能观测到的随机序列，都可以看成是确定性的系统的实现. 它们呈现的“随机性态”是由于系统的初值灵敏性和系统的瞬变性等混沌性质所造成. 如

何将以上的不同观点统一在同一个概念范畴里，将是本节讨论的主要内容.

在第4章§4.6中，曾引入过时间序列的第二种可逆性，即时间可逆性，并用了若干例子说明纯随机平稳序列与单向随机平稳序列等概念. 从单向随机序列的例4.6.7已经看到，混沌与随机序列可以统一在同一个平稳序列中. 在这一小节中，我们将进一步阐述它们的内在联系.

先从确定量与随机量的本质区别来说起. 以 x 表示某个随机变量，如果 $P(x=c)=1$，即 x 取 c 的概率为1，可称此量 $x=c$ 为确定量，又可称 x 为在 c 点退化的随机变量. 反之，如果对任何常数 c，$P(x=c)<1$，则称 x 为非退化的随机变量，简称为非退化的. 此时 x 的确具有一定的随机性态. 由此可见，确定量与非确定的随机变量可以用概率分布统一在随机变量的概念中，并用随机变量分布的退化与非退化区分它们的本质差别.

对于二维随机向量 $\boldsymbol{x}=(x_1,x_2)^\tau$，仍可用其分布的退化与非退化来区分确定量与非确定量. 但是，二维分布的退化现象不仅包括了随机向量退化为常数向量，还包括每个分量的退化，以及两个分量之间的依赖关系的退化，即指两个分量有确定的依赖关系. 前两种退化问题容易从一维情况得到推广性的定义，例如当 $P(\boldsymbol{x}=\boldsymbol{c})=1$ 时，称 $\boldsymbol{x}$ 为退化随机向量；当 $P(x_i=c_i)=1\ (i=1$ 或 $2)$时，称 $\boldsymbol{x}$ 的第 i 分量为退化的. 但是，前述的第三种退化现象要借助于条件分布来描述. 在一维情况并不存在此类退化，为此用 $F(x,y)$表示 $\boldsymbol{x}$ 的二维联合分布，并记

$$F_1(x|y)=P(x_1<x|x_2=y), \tag{8.1.12}$$

$$F_2(x|y)=P(x_2<x|x_1=y), \tag{8.1.13}$$

分别为给定 $x_2=y$ 时 x_1 的条件分布，及给定 $x_1=y$ 时 x_2 的条件分布. 于是，按照上述的分布为退化与非退化进行分类时，可将条件分布分成如下的四种类型：

(i) $F_1(x|y)$和 $F_2(x|y)$都是完全退化的，这里称 $F_i(x|y)$为完全退化的，是指对每个 y 而言，$F_i(x|y)$为退化的条件分布；

(ii) $F_1(x|y)$ 是完全退化的，而 $F_2(x|y)$ 不是；

(iii) $F_2(x|y)$ 是完全退化的，而 $F_1(x|y)$ 不是；

(iv) $F_1(x|y)$ 和 $F_2(x|y)$ 都不是完全退化的.

易见，以上四类又依次等价于如下的四类：

(i′) 存在函数 φ 和 ψ，使得

$x_1=\varphi(x_2)$, $x_2=\psi(x_1)$;

(ii′) 存在函数 φ,使得

$x_1=\varphi(x_2)$,但不存在(i′)中的 ψ;

(iii′) 存在函数 ψ,使得

$x_2=\psi(x_1)$,但不存在(i′)中的 φ;

(iv′) 不存在(i′)中的 φ 和 ψ.

例 8.1.14 若 $\boldsymbol{x}=(x_1,x_2)^\tau$,其中 x_1 和 x_2 为相互独立的随机变量,而且都不是退化的,显然 $\boldsymbol{x}$ 的条件分布属于第(iv)类.

例 8.1.15 若 $\boldsymbol{x}=(x_1,x_1)^\tau$,其中 x_1 为非退化的随机变量,易见此时 $\boldsymbol{x}$ 的条件分布属于第(i)类,且 $\varphi(x)=\psi(x)=x$.

例 8.1.16 若 $\boldsymbol{x}=(x_1,x_1^2)^\tau$,其中 x_1 为非退化的随机变量,不难看出,当 $P(x_1\geqslant 0)=1$ 或者 $P(x_1\leqslant 0)=1$ 时,$\boldsymbol{x}$ 的条件分布属于第(i)类. 当 $0<P(x_1\geqslant 0)<1$ 时,$\boldsymbol{x}$ 的条件分布属于第(iii)类,因为 $x_2=x_1^2$,即 $\psi(x_1)=x_1^2$,而 $x_1=\pm\sqrt{x_2}$ 不能被 x_2 唯一确定.

例 8.1.17 若 $\boldsymbol{x}=(1,x_2)^\tau$,其中 x_2 为非退化的随机变量,显然 $\boldsymbol{x}$ 的条件分布属于第(ii)类,而且 $\varphi(x)=1$,即 $x_1=\varphi(x_2)=1$.

例 8.1.18 若 $\boldsymbol{x}=(u,u/2+\varepsilon/2)^\tau$,其中 u 与 ε 为相互独立的随机变量. 而且 $x_1=u$ 在[0,1]上均匀分布 $P(\varepsilon=0)=P(\varepsilon=1)=\frac{1}{2}$,$x_2=u/2+\varepsilon/2$,其分布也是[0,1]上的均匀分布. 若记 $\varphi(x)=2x\bmod(1)$,容易验证

$$x_1=u=\varphi(u/2+\varepsilon/2)=\varphi(x_2).$$

由此可知,$\boldsymbol{x}$ 的条件分布属于第(ii)类. 但是,如果以上的 ε 不是二项分布,而是有正密度的随机变量,则上述的 $\boldsymbol{x}$ 的条件分布属于第(iv)类.

例 8.1.19 若 $\boldsymbol{x}=(x_1,x_2)^\tau$ 为二维正态随机变量,其均值为零向量,协方差阵为

$$\Gamma=\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}.$$

于是 $\boldsymbol{x}$ 的联合分布密度为

$$f(x,y)=\frac{1}{\sqrt{2\pi(1-\rho)^2}}e^{-\frac{1}{2(1-\rho^2)}(x^2+y^2-2\rho xy)}.$$

如果 $x_2=x_1$,即 $\boldsymbol{x}=(x_1,x_1)^\tau$ 的条件分布为退化的第(i)类情况,显然此时

$\rho=1$. 这时,上式的 $f(x,y)$ 不再有意义,也就是说,$\boldsymbol{x}$ 不再有二维分布密度.

在一般情况下(不一定是正态分布),只要 $\boldsymbol{x}=(x_1,x_2)^{\tau}$ 的条件分布属于(i)、(ii)或(iii)类,$\boldsymbol{x}$ 都不存在二维联合密度函数. 反之,若 $\boldsymbol{x}$ 没有二维联合密度,并不意味着 $\boldsymbol{x}$ 的条件分布不属于第(iv)类. 例如在例 8.1.14 中的 x_1 和 x_2 都是离散分布时,$\boldsymbol{x}=(x_1,x_2)^{\tau}$ 的条件分布属于第(iv)类,但是它也没有联合密度.

以上关于条件分布的退化分类的讨论,无疑可以推广到多维随机变量的情况,只是分类更繁琐一些,并无实质性的困难和新意. 进一步还可推广到无穷维的随机序列的情况. 考虑时间序列 $\{x_t,t\geqslant 0\}$,如果它的任何有限维联合分布的条件分布都不退化,则它属于第 4 章中定义4.6.10 的纯随机平稳序列. 如果 $\{x_t,t\geqslant 0\}$ 是定义 4.6.11 的单向随机平稳序列,那么它是一种特殊的条件分布有退化现象的时间序列. 对于这种单向随机序列,向前看时,它是确定性的依赖关系;向后看时,它是非确定性的随机依赖关系. 或者说,向前看是确定性序列,向后看则是随机序列. 结合定义 4.6.11 和条件分布的退化的分类的讨论,便能理解由混沌变换产生的迭代序列为什么像随机序列. 其实,反方向看它们时,它们本来就是在严格概率意义下的随机序列.

以上所讨论的内容,从时间序列的条件分布的退化性质揭示了时间序列与混沌之间的联系. 它们被统一在随机序列的概念中,并以纯随机或单向随机将它们的本质差别区分开. 另外,如同在第 2 章和第 4 章中所提到的那样,时间序列模型与确定的动态系统存在着密切联系,如自回归模型与其骨架模型的联系即属于此类联系. 因此,这也从另一方面表现出时间序列与混沌之间的联系. 以 p 阶非线性自回归模型为例,就有如下形式:

$$x_t=\varphi(x_{t-1},\cdots,x_{t-p})+\varepsilon_t. \tag{8.1.14}$$

其向量形式如(4.2.29)式,即

$$\begin{cases}\boldsymbol{x}_t=\boldsymbol{\Phi}(\boldsymbol{x}_{t-1})+\varepsilon_t\boldsymbol{u} \quad (t\geqslant 1);\\ \boldsymbol{x}_0\in R^p.\end{cases} \tag{8.1.15}$$

式中 $\boldsymbol{u}=(1,0,\cdots,0)^{\tau}$,$\boldsymbol{x}_t=(x_t,x_{t-1},\cdots,x_{t-p+1})^{\tau}$,

$$\boldsymbol{\Phi}(\boldsymbol{x}_{t-1})=(\varphi(\boldsymbol{x}_{t-1}),x_{t-1},\cdots,x_{t-p+1})^{\tau}.$$

与(8.1.14)式相应的骨架模型是

$$y_t=\varphi(y_{t-1},\cdots,y_{t-p}) \quad (t\geqslant 1). \tag{8.1.16}$$

与(8.1.15)式相应的骨架模型是

$$\begin{cases} \boldsymbol{y}_t = \boldsymbol{\Phi}(\boldsymbol{y}_{t-1}) \quad (t \geqslant 1); \\ \boldsymbol{y}_0 \in R^p. \end{cases} \tag{8.1.17}$$

在第 4 章中我们曾看到,自回归模型(8.1.15)的平稳解和遍历性的条件与动态系统(8.1.17)的稳定性有密切关系.我们曾形容过,模型(8.1.15)是确定性动态系统(8.1.17)披上干扰项 $\varepsilon_t\boldsymbol{u}$ 外衣而得到的.当 φ 为非线性函数时,(8.1.15)及(8.1.14)式为非线性时间序列模型,而相应的(8.1.17)或(8.1.16)式为非线性动态系统,它们便会卷入混沌概念.时间序列与混沌的这种联系表现得更直接,而且有着广泛的实际背景.由于任何确定性的动态系统难免卷入各种干扰项或者卷入观测误差项,这都使得系统(8.1.17)常被(8.1.15)式所代替.即使被观测到的数据序列果真满足(8.1.17)式,作为非线性时间序列分析,也会关心数据序列是否满足(8.1.17)式的统计判别问题.总之,研究非线性时间序列分析与研究混沌有着密切的内在联系.

最后顺便指出:第 4 章中讨论的马氏链的遍历性和时间序列模型的遍历性,与本节讨论的确定系统的不变分布的遍历性以及文献中随机过程的遍历性,是几种不相同的遍历概念.它们或许有某种联系,但与本书内容无关,这里不予讨论了.

§ 8.2　经验分布与伪随机序列

在这一节中,我们将通过对序列的经验分布的讨论,说明时间序列的统计分析也与混沌序列的统计分析有联系.还要介绍一种特殊的混沌序列,即伪随机序列以及它的经验分布的收敛性质.

如果 $x_1, x_2, \cdots, x_n$ 是某个平稳序列 $\{x_t\}$ 的样本序列,基于此样本序列的统计方法和统计理论,在文献中已有丰富的内容.其中最基本的统计量是如下的样本分布,又称经验分布:

$$F_n(x) = \frac{1}{n}\sum_{t=1}^{n} I(x_t < x) \quad (-\infty < x < \infty), \tag{8.2.1}$$

式中 $I(x_t < x)$ 表示示性函数.在适当的条件下,基于大数定律和平稳性有如下的熟知结果,即

$$\lim_{n\to\infty} F_n(x) = F(x), \text{ a.s.}, \tag{8.2.2}$$

其中 F 为 x_t 的分布函数.这一性质被称为样本分布 F_n 的相容性.

如果 $x_1,x_2,\cdots,x_n$ 是满足形如(8.1.1)式的确定系统的迭代序列，在实际中，系统的变换 φ 和初始值都可能是未知的，则在形式上仍可用(8.2.2)式定义此数据序列的样本分布 F_n. 但是，当 n 趋于无穷大时，F_n 是否有极限，在什么意义下有极限，其极限是否为分布函数，以及极限函数有什么意义，这些都是在统计意义下最为关心的问题，也是为实际数据分析所关心的问题. 然而，当 $x_1,x_2,\cdots,x_n$ 不是平稳序列的样本时，时间序列分析中的理论与技巧就不再奏效了. 所以研究确定性数据序列的经验分布时，就需要寻找适当的理论与方法.

在这一节中，我们将考虑两种不同类型的确定性序列来讨论它们的经验分布的极限问题. 其中一类是属于前面提到的单向随机平稳序列；另一类是正反方向都为确定性序列，但又有某种意义下的混沌特征，即所谓的伪随机序列，其详细定义在以后给出.

首先讨论单向随机平稳序列的情况，并且以例 8.1.2 的动态系统为例，即

$$\begin{cases} x_t = 2x_{t-1}\bmod(1) \quad (t \geqslant 1); \\ x_0 \in [0,1]. \end{cases} \tag{8.2.3}$$

由 § 4.6 的讨论知，系统(8.2.3)的迭代序列为单向随机平稳序列，而且此序列的反向序列满足如下的 LAR(1)模型，即令 $z_t=x_{n-t}$时，它满足

$$z_t = \frac{1}{2}z_{t-1} + \frac{1}{2}\varepsilon_t \quad (t = 1,2,\cdots,n-1), \tag{8.2.4}$$

其中$\{\varepsilon_t\}$为独立同分布序列，而且 $P(\varepsilon_t=0)=p, P(\varepsilon_t=1)=1-p$，$\varepsilon_t$ 与 $\{z_s, s\leqslant t-1\}$相互独立. 但是，在此必须指出：当 n 变大时，不仅 $z_1,z_2,\cdots,z_n$ 的数据个数增加，而且其各元素也随 n 而改变. 所以应将(8.2.4)式改写为

$$z_{nt} = x_{n-t} \quad (t = 0,1,\cdots,n-1), \tag{8.2.5}$$

$$z_{nt} = \frac{1}{2}z_{n,t-1} + \frac{1}{2}\varepsilon_{nt} \quad (t = 1,2,\cdots,n-1), \tag{8.2.6}$$

其中$\{\varepsilon_{nt}\}$在 n 固定时是独立同分布序列，且 ε_{nt}与$\{z_{ns}, s<t\}$独立，ε_{nt}与 ε_t 的分布相同. 实际上，序列$\{z_{nt}, t=1,2,\cdots,n, n\geqslant 1\}$被称为*平稳序列的双重排列*. 也就是说，对每个固定的 n，$\{z_{n1},z_{n2},\cdots,z_{nn}\}$与满足(8.2.4)式的$\{z_1,z_2,\cdots,z_n\}$具有完全相同的概率分布，前者是随 n 增大重新排列的；而后者是单调增长的序列.

由(8.2.5)式知，$x_1,x_2,\cdots,x_n$ 的经验分布 F_n 同时又是模型(8.2.6)

式的样本的经验分布，即

$$F_n(x) = \frac{1}{n}\sum_{t=1}^{n} I(x_t < x) = \frac{1}{n}\sum_{t=1}^{n} I(z_{nt} < x). \qquad (8.2.7)$$

现在考虑与系统(8.2.3)相应的自回归模型(8.2.4)中的 p 值选择问题. 从模型(8.2.4)看，p 可取(0,1)中任何值. 从(8.2.7)式左端等式看，经验分布 F_n 并不依赖于 p 值. 这便产生了疑问：不同的 p 值选择将导致什么样的结果？有没有一个特殊的 p 值具有某种特殊的意义？这些问题与经验分布 F_n 的极限行为有关. 而 F_n 的极限行为非常复杂，在此不能详加讨论，下面仅叙述其某些重要特征，主要为了介绍与时间序列分析的联系.

首先，根据前一节的讨论可知，[0,1]上的均匀分布是系统(8.2.3)的不变分布，而且是遍历的. 再用性质 8.1.13 并取 $g(x) = I(x_1 < x)$，由(8.1.10)式知

$$\lim_{n\to\infty} F_n(x) = \int_0^x \mathrm{d}y = x \quad (0 \leqslant x \leqslant 1), \text{a.s. } x_0,$$

为了叙述方便，我们用 $U(x)$ 表示[0,1]上的均匀分布，即

$$U(x) = \begin{cases} 0, & \text{当 } x < 0; \\ x, & \text{当 } 0 \leqslant x \leqslant 1; \\ 1, & \text{当 } x \geqslant 1. \end{cases}$$

则前一式可以写成

$$\lim_{n\to\infty} F_n(x) = U(x), \text{ a.s. } x_0. \qquad (8.2.8)$$

另一方面，利用(8.2.5)、(8.2.6)和(8.2.7)式，并取 $p = \frac{1}{2}$，按时间序列分析方面考虑后的极限问题，在例 4.2.12 中曾论证过，此时模型(8.2.4)以及(8.2.6)的平稳分布为[0,1]上的均匀分布 $U(x)$，只是由于(8.2.7)式右端等式表明了 F_n 是双重排列的$\{z_{nt}, t = 1,2,\cdots,n\}$的经验分布，从而不能直接使用(8.2.2)式. 但是，利用(8.2.6)式和 § 1.5 中的向后推移法可得到

$$z_{nt} = \frac{1}{2}\varepsilon_{nt} + \left(\frac{1}{2}\right)^2 \varepsilon_{n,t-1} + \cdots + \left(\frac{1}{2}\right)^t \varepsilon_1 + \left(\frac{1}{2}\right)^{t+1} z_{n0}, \quad (8.2.9)$$

再由(8.2.7)式右端等式知

$$E\left\{\frac{1}{n}\sum_{t=1}^{n} I(z_{nt} < x)\right\} = \frac{1}{n}\sum_{t=1}^{n} EI(z_{nt} < x)$$

$$= \frac{1}{n}\sum_{t=1}^{n} V(x) = U(x). \qquad (8.2.10)$$

联合以上两式,并用概率方法计算 $E[F_n(x)-U(x)]^4$,可以得到如下的阶数估计,即

$$E\{F_n(x) - U(x)\}^4 = E\left\{\frac{1}{n}\sum_{t=1}^{n}[I(z_{nt} < x) - x]\right\}^4$$

$$= O\left(\frac{1}{n^2}\right) \quad (0 \leqslant x \leqslant 1). \qquad (8.2.11)$$

再用 Borel Cantelli 引理可得

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} I(z_{nt} < x) = V(x), \text{ a.s.} \qquad (8.2.12)$$

以上详细论证,都是时间序列分析中的典型方法,这里从略. 对比(8.2.12)和(8.2.8)式可知,当 $p=\frac{1}{2}$时,用纯概率方法也能得到相同的结果. 而(8.2.8)式的证明则使用了动态系统的不变分布和遍历分布的概念,它们可以不涉及概率统计的概念. 通过此例分析可看出,时间序列分析的理论与方法可以用于研究确定性序列的统计分析,至少对单向随机的确定性序列是如此.

关于系统(8.2.3)的迭代序列的经验分布的极限问题,还有两点值得指出:其一,并不是对任何初值 x_0 而言 F_n 都有极限. 但是,使 F_n 没有极限的初值 x_0 的全体也只能是 Lebesgue 零测集合. 其二,除了均匀分布以外,F_n 的确还有其他极限,同样地,根据(8.2.8)式知,使 F_n 收敛于其他分布的初值 x_0 的全体也只是 Lebesgue 零测集. 为了对此两点作一简要说明,我们引用 x_0 的二进制的表示形式,即

$$x_0 = 0.a_0a_1a_2\cdots,$$

其中 $a_0,a_1,\cdots$表示只取 0 或 1 的符号,显然,$x_0\in[0,1]$. 根据(8.2.3)式可见

$$x_t = 0.a_ta_{t+1}a_{t+2}\cdots \quad (t = 0,1,2,\cdots), \qquad (8.2.13)$$

其中 x_t 也是用二进制所表示的小数. 容易验证如下的事实:

(i) 当 $x_0=0.a_0a_1\cdots a_m$ 时,即 x_0 为有限的二进制小数时,则当 $t>m$ 时,$x_t=0$,从而易见

$$\lim_{n\to\infty} F_n(x) = J_0(x). \qquad (8.2.14)$$

式中 J_0 表示在 0 点退化的分布函数,即

$$J_0(x) = 0,\ 当\ x \leqslant 0;\ J_0(x) = 1,\ 当\ x > 0.$$

(ii) 当 $x_0=0.a_0a_1\cdots a_m\dot{a}_{m+1}\cdots\dot{a}_{m+q}$，即 x_0 在第 $m+1$ 位开始为循环小数，于是由(8.2.3)式有

$$x_{m+kq+i}=x_{m+kq+q+j}\quad(k\geqslant 0;j=1,2,\cdots,q).$$

也就是说，$\{x_t,t\geqslant 0\}$ 的尾部序列是周期的，而且以 q 为其周期长度，从而易见

$$\lim_{n\to\infty}F_n(x)=J_q(x). \tag{8.2.15}$$

式中 J_q 表示在以下 q 个点取相同的概率 $1/q$ 的跳跃分布，此 q 个点为：

$$c_1=0.a_{m+1}a_{m+2}\cdots a_{m+q-1}a_{m+q},$$
$$c_2=0.a_{m+2}a_{m+3}\cdots a_{m+q}a_{m+1},$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$c_q=0.a_{m+q}a_{m+1}\cdots a_{m+q-2}a_{m+q-1}.$$

在此还须指出：在(i)中，当 x_0 为有限二进制小数时，必有(8.2.14)式成立. 但是，反之不然，例如取

$$x_0=0.0110100010\cdots010\cdots, \tag{8.2.16}$$

其中 $a_0=0$，$a_i=0\quad(i\neq 2^j,j=0,1,\cdots)$，其余 $a_i\neq 0$，根据(8.2.1)式不难验证(8.2.14)式仍成立. 类似地，满足(8.2.15)式的初值 x_0 也不一定不是循环二进小数. 最后顺便指出，利用二进制表示方法，不难给出特殊的 x_0，使得 F_n 没有极限，读者可以自己尝试去找到这样的 x_0 值. 但是，要找到一 x_0 值的例子，证明相应的(8.2.8)式成立却不是容易的事，虽然(8.2.8)式对几乎所有的 $x_0\in[0,1]$均成立.

以上讨论了系统(8.2.3)的样本序列的经验分布的极限问题. 然而(8.2.3)仅仅是一个很简单的非线性系统. 说它很简单，首先是因为(8.2.3)中的非线性函数 $\varphi(x)=2x\bmod(1)$ 比较简单，更特别的是：(8.2.3)式的反向序列模型是一种最简单的一阶线性自回归模型. 一般说来，非线性平稳序列的反向序列模型不一定是线性时间序列模型. 这可参见§4.6所举的几个例子. 因此，上面所讨论的(8.2.3)式的经验分布的收敛问题推广到更一般的动态系统仍是比较困难的问题.

下面再讨论另一种确定性序列，它既非纯随机序列，也不是单向随机序列. 如果不加任何其他限制，这样的序列太多了，也没有必要研究它们的经验分布. 例如 $x_t=a+bt\quad(t=0,1,\cdots)$ 就是这样的例子，它的经验分布没有研究的价值. 所以说，我们要讨论的确定性序列仍然是某个动态系统的轨道序列. 确切地说，它们是满足如下定义的伪随机序列.

定义 8.2.1　系统(8.1.1)的迭代序列$\{x_t, t\geqslant 0\}$被称为伪随机序列，如果：

(i) φ是瞬变的；

(ii) φ有遍历的不变分布F，F与$[0,1]$上的 Lebesgue 测度绝对连续；

(iii) 存在函数ψ，使得$x_t=\psi(x_{t+1})$　$(t=0,1,\cdots)$.

在以上定义中，根据(i)、(ii)和定义 8.1.12，系统(8.1.1)中的φ是混沌变换. 根据(iii)可知，当$x_t=\varphi_t(x_0)$满足条件(iii)时，必然存在φ的逆函数$\varphi^{-1}=\psi$，使得$x_t=\psi(x_{t+1})$. 于是，对于上述的伪随机序列而言，无论正看还是反看，此序列都是确定性的. 由于φ又是依定义 8.1.12 为混沌的，迭代序列$\{x_t, t\geqslant 0\}$又呈现一定的随机性态，故称它为伪随机序列.

根据定义 8.2.1 还可知道由系统(8.1.1)迭代的序列，不一定都是伪随机序列，而只对某些初值x_0而言才是伪随机序列. 再据条件(i)、(ii)和性质(8.2.2)式可知，对于几乎所有的伪随机序列，其经验分布都满足(8.2.2)式. 为了说明上述伪随机序列的存在性，特举出如下的例子：

例 8.2.2　考虑§8.1的动态系统(8.1.1)，其中变换φ由下式定义：

$$\varphi(x)=\begin{cases}\dfrac{3}{2}x+\dfrac{1}{4}, & \text{当 } 0\leqslant x\leqslant\dfrac{1}{2};\\[2mm] \dfrac{1}{2}x-\dfrac{1}{4}, & \text{当 } \dfrac{1}{2}<x\leqslant 1.\end{cases}\tag{8.2.17}$$

我们将证明以上的φ满足定义 8.2.1 中的(i)和(ii). 而且易见φ在$(0,1)$上有逆函数φ^{-1}，且为如下形式：

$$\varphi^{-1}(y)=\begin{cases}2y+\dfrac{1}{2}, & \text{当 } 0<y\leqslant\dfrac{1}{4};\\[2mm] \dfrac{2}{3}y-\dfrac{1}{6}, & \text{当 } \dfrac{1}{4}<y<1.\end{cases}$$

由此可见，对每个$x_0\in(0,1)$而言，由(8.2.17)式中的φ迭代的序列都是伪随机序列. 下面分别证明(i)和(ii). 为此，引进如下的代换.

首先采用代换$y_t=x_t+\dfrac{1}{2}$，于是由(8.2.17)和(8.1.1)决定的动态系统可得到y_t所满足的动态系统为

$$\begin{cases}y_t=g(y_{t-1})\quad(t\geqslant 1);\\[2mm] y_0\in\left[\dfrac{1}{2},\dfrac{3}{2}\right].\end{cases}\tag{8.2.18}$$

其中

$$g(y)=\begin{cases}\frac{3}{2}y, & \text{当}\frac{1}{2}\leqslant y\leqslant 1;\\ \frac{1}{2}y, & \text{当}1<y\leqslant\frac{3}{2}.\end{cases}$$

再用代换 $z_t=\log y_t$，由(8.2.18)式可得到 y_t 所满足的动态系统为

$$\begin{cases}z_t=h(z_{t-1}) \quad (t\geqslant 1);\\ z_0\in\left[\log\frac{1}{2},\log\frac{3}{2}\right],\end{cases} \tag{8.2.19}$$

其中

$$h(z)=\begin{cases}z+\log\frac{3}{2}, & \text{当}\log\frac{1}{2}\leqslant z\leqslant 0;\\ z+\log\frac{1}{2}, & \text{当}0<z\leqslant\log\frac{3}{2}.\end{cases}$$

再用代换 $u_t=(z_t+\log 2)/\log 3$，由(8.2.19)式可得

$$\begin{cases}u_t=\xi(u_{t-1}) \quad (t\geqslant 1);\\ u_0\in[0,1],\end{cases} \tag{8.2.20}$$

其中

$$\xi(u)=\begin{cases}u+\alpha, & \text{当}0\leqslant u\leqslant 1-\alpha;\\ u+\alpha-1, & \text{当}1-\alpha<u\leqslant 1.\end{cases}$$

式中 $\alpha=(\log 3-\log 2)/\log 3$，显然这是一个无理数.上式又可写成

$$\xi(u)=(u+\alpha)\mathrm{mod}(1).$$

于是(8.2.20)式可写成

$$\begin{cases}u_t=(u_{t-1}+\alpha)\mathrm{mod}(1) \quad (t\geqslant 1);\\ u_0\in[0,1].\end{cases} \tag{8.2.21}$$

最后再用代换 $w_t=e^{i2\pi u_t}$，则由(8.2.21)式可得到

$$\begin{cases}w_t=e^{i2\pi\alpha}.w_{t-1} \quad (t\geqslant 1);\\ w_0\in C\equiv\{e^{i2\pi\theta},0<\theta\leqslant 1\}.\end{cases} \tag{8.2.22}$$

令 $T(w)=e^{i2\pi\alpha}\cdot w$，可知(8.2.22)式是将复平面上的单位圆周 C 进行角度为 $2\pi\alpha$ 的旋转变换.因此，圆周 C 上的 Lebesgue 测度是其不变测度.又由于(8.2.22)中的 α 为无理数，根据文献[117]中的定理 1.8 知，旋转变换 T 是瞬变的，而且 C 上的 Lebesgue 测度是遍历的.根据(8.2.22)与(8.2.21)的联系可知，系统(8.2.21)式的遍历的不变分布是[0,1]上的均匀分布.然后，依次利用系统(8.2.21)、(8.2.20)、(8.2.19)、(8.2.18)和(8.2.17)式的联系，便可验证(8.2.17)式的不变分布

$$F(x)=\left\{\log\left(x+\frac{1}{2}\right)+\log 2\right\}\Big/\log 3 \quad (0\leqslant x\leqslant 1) \qquad (8.2.23)$$

是遍历的，而且显然满足定义 8.2.1 中的第(ii)条.

利用上面论述的结果以及(8.2.23)和(8.2.2)式，由(8.2.17)式中的 φ 所产生的迭代序列 $x_0,x_1,\cdots,x_n$，其经验分布 F_n 必满足：对于 $0\leqslant x\leqslant 1$，有

$$\lim_{n\to\infty}F_n(x)=\left\{\log\left(x+\frac{1}{2}\right)+\log 2\right\}\Big/\log 3,\ \text{a.s. } x_0. \quad (8.2.24)$$

由于(8.2.23)式的分布在[0,1]上与 Lebesgue 测度相互绝对连续，所以(8.2.24)式中的“a.s. x_0”是对几乎所有的 Lebesgue 测度意义下的 $x_0\in[0,1]$而言的. 那么到底使(8.2.24)式不成立的 x_0 有哪些呢？一般说来，很难回答这一问题. 例如在讨论系统(8.2.3)时，曾指出有(8.2.14)和(8.2.15)式出现，而且对 x_0 为[0,1]中的某些二进制无理数能使(8.2.8)式的极限也不成立. 其实我们还没有详细讨论使(8.2.8)式不成立的所有 x_0 的情况. 对于系统(8.2.22)而言，利用旋转性质不难验证对于任何初值 w_0，由(8.2.22)式迭代的序列$\{w_t,t\geqslant 0\}$的相应的幅角序列为$\{2\pi u_t, t\geqslant 0\}$，其中 $0\leqslant u_t\leqslant 1$ 满足(8.2.21)式；并且$\{u_t,t\geqslant 0\}$的经验分布总满足(8.2.2)式，式中 $F(x)=U(x)$. 再次利用(8.2.17)式～(8.2.22)式的联系可知，(8.2.24)式对每个 x_0 均成立. 换言之，使(8.2.24)式不成立的初值 x_0 的全体不仅是 Lebesgue 零测集，而且是空集. 对于伪随机序列而言，这是一个有意义的性质，它表明了无论从怎样的 $x_0\in(0,1)$出发，由(8.2.17)式的 φ 所迭代出的伪随机序列，其经验分布总使(8.2.24)式成立.

关于伪随机序列的数值例子，可参见例 2.2.8.

§8.3 非线性时间序列与分形维数

在研究非线性动态系统时，其吸引集、吸引域的几何特征也表现出惊人的复杂性，如见例 2.3.25 所示. 研究这些吸引集、吸引域的几何结构属于近代的分形几何学的内容. 分形几何与动态系统密切相关，也是近年来比较受关注的研究领域，如见文献[87]. 要详细掌握具体非线性动态系统的吸引集和吸引域的分形几何结构，即使对本书中的离散情况，一般说来，也是非常困难的. 目前利用推广的维数概念刻画分形几何中的各类集

合特征，是被研究较多的内容. 这些分形维数能表达复杂集合的部分特征. 它们的定义和估计方法不同程度地与时间序列分析有联系. 本节将给出简要介绍.

一、Lyapunov 指数与维数

在讨论 Lyapunov 指数与维数时，我们仅限于考虑动态系统中的 φ 为连续可微的情况. 仍然从最简单的一元一阶线性变换说起，即

$$\varphi(x) = ax + b.$$

重复迭代以上 φ 变换，可得到

$$\varphi_n(x)=a^n x+(a^{n-1}+a^{n-2}+\cdots+a+1)b,\ n\geqslant 1. \tag{8.3.1}$$

将上式中 φ 的 $x_0=x$ 换以 $x_0=y$，并求 $\varphi_n(x)$ 与 $\varphi_n(y)$ 的差，可得到

$$\varphi_n(x) - \varphi_n(y) = a^n(x-y) = \frac{\mathrm{d}}{\mathrm{d}u}\varphi_n(u)(x-y), \tag{8.3.2}$$

式中

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}u}\varphi_n(u) &= \frac{\mathrm{d}}{\mathrm{d}u}\varphi(\varphi_{n-1}(u)) = \frac{\mathrm{d}}{\mathrm{d}u}(a\varphi_{n-1}(u)+b) = a\frac{\mathrm{d}}{\mathrm{d}u}\varphi_{n-1}(u)\\ &= a^2\frac{\mathrm{d}}{\mathrm{d}u}\varphi_{n-2}(u) = \cdots = a^{n-1}\frac{\mathrm{d}}{\mathrm{d}u}(au+b) = a^n.\end{aligned} \tag{8.3.3}$$

由(8.3.2)式又有

$$\frac{1}{n}\log\left|\frac{\mathrm{d}}{\mathrm{d}u}\varphi_n(u)\right| = \log|a| \equiv \lambda. \tag{8.3.4}$$

显然，当 $\lambda<0$ 时，即 $|a|<1$，(8.3.1)式的迭代序列是收敛的；当 $\lambda>0$ 时，$\varphi_n(x)$ 是发散的. 这与第 2 章 § 2.1 中所讨论的(2.1.2)式相似.

稍广泛一些，考虑 φ 为一元连续可微的非线性变换，则(8.3.1)式被如下的表示所代替，即

$$\varphi_n(x) = \varphi(\varphi_{n-1}(x)) = \cdots = \varphi(\varphi(\cdots\varphi(x)\cdots)). \tag{8.3.5}$$

同时，(8.3.2)式可用如下的近似等式代替：

$$\varphi_n(x) - \varphi_n(y) \approx \frac{\mathrm{d}}{\mathrm{d}u}\varphi_n(u)\Big|_{u=x}(x-y). \tag{8.3.6}$$

此时(8.3.3)式应为

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}u}\varphi_n(u) &= \frac{\mathrm{d}}{\mathrm{d}u}\varphi(\varphi_{n-1}(u)) = \dot\varphi(\varphi_{n-1}(u))\frac{\mathrm{d}}{\mathrm{d}u}\varphi_{n-1}(u)\\ &= \dot\varphi(\varphi_{n-1}(u))\cdot\dot\varphi(\varphi_{n-2}(u))\cdots\dot\varphi(u),\end{aligned}$$

式中 $\dot{\varphi}(u)=\frac{\mathrm{d}}{\mathrm{d}u}\varphi(u)$. 记 $x_k=\varphi_k(x)$, $x_0=x$, 于是前一式为

$$\frac{\mathrm{d}}{\mathrm{d}u}\varphi_n(x)=\prod_{k=0}^{n-1}\dot{\varphi}(x_k)=\prod_{k=0}^{n-1}\frac{\mathrm{d}}{\mathrm{d}u}\varphi(u)|_{u=x_k}. \tag{8.3.7}$$

类似地,(8.3.4)式被下式代替:

$$\lambda(x)=\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\log|\dot{\varphi}(x_k)|. \tag{8.3.8}$$

依上式定义 $\lambda(x)$ 时,存在两个问题:一是右端的极限是否存在,二是极限值是否依赖于初始值 $x_0=x$. 当 φ 为线性函数时,如(8.3.4)式所示,(8.3.8)式右端的极限存在,而且与初值 x_0 无关. 在 φ 为非线性函数时,如果 φ 不仅连续可微,而且满足定义 8.1.12 的条件(i)和(ii),根据性质 8.1.13 知,(8.3.8)式的极限存在,且与初值 x_0 无关,至少在几乎处处的意义下是这样的. 显而易见,这时由(8.3.8)式定义的 $\lambda(x)=\lambda$ 是变换 φ 的重要特征描述. 这样的 λ 被称为 φ 的 Lyapunov 指数.

以上叙述可以推广到从 R^m 到 R^m 的变换情况. 考虑 m 维向量 $\boldsymbol{x}=(x_1,\cdots,x_m)^\tau$ 的变换

$$\boldsymbol{\Phi}(\boldsymbol{x})=(\varphi_1(\boldsymbol{x}),\varphi_2(\boldsymbol{x}),\cdots,\varphi_m(\boldsymbol{x}))^\tau.$$

这里要求变换有各分量的一阶偏导数. 相应于(8.3.7)式的推广为

$$\begin{aligned}D\boldsymbol{\Phi}_n(\boldsymbol{x})&=\frac{\partial}{\partial\boldsymbol{x}}\boldsymbol{\Phi}_n(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{x}_0}=\prod_{t=0}^{n-1}\frac{\partial}{\partial\boldsymbol{x}}\boldsymbol{\Phi}(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{x}_0}\\&=\prod_{t=0}^{n-1}\boldsymbol{\Phi}'(\boldsymbol{x}_t),\end{aligned} \tag{8.3.9}$$

式中

$$\boldsymbol{\Phi}'(\boldsymbol{x})=\frac{\partial}{\partial\boldsymbol{x}}\boldsymbol{\Phi}(\boldsymbol{x})=\left(\frac{\mathrm{d}}{\mathrm{d}x_i}\varphi_j(\boldsymbol{x})\right)_{1\leqslant i,j\leqslant m}.$$

为了给出(8.3.8)式的推广形式,先求出 Jacobi 矩阵 $D\boldsymbol{\Phi}_n(\boldsymbol{x})$ 特征值的模数,并按其大小排列为

$$a_{n1}(\boldsymbol{x})\geqslant a_{n2}(\boldsymbol{x})\geqslant\cdots\geqslant a_{nm}(\boldsymbol{x})\geqslant 0.$$

于是类似于(8.3.8)式可定义 λ_i 为

$$\lambda_i=\lambda_i(\boldsymbol{x})=\lim_{n\to\infty}\frac{1}{n}\log a_{ni}(\boldsymbol{x})\quad(i=1,2,\cdots,m). \tag{8.3.10}$$

与(8.3.8)式类似,在适当的条件下,(8.3.10)式的极限存在,且与初始向量 $\boldsymbol{x}_0$ 无关,至少在几乎处处意义下是正确的. 由于本书未介绍多维情况下的遍历概念,这里不讨论(8.3.10)式成立的条件和例子. 有兴趣的读者

可参见文献[106]. 由(8.3.10)式定义的 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_m$ 被称为变换 Φ 的 Lyapunov 指数. 这些指数的大小反映了由变换 Φ 迭代产生的序列的特征. 例如若某个 $\lambda_i < 0$, 则反映了在某个方向上迭代序列收缩的平均性态. 又如, 当 Φ 为多维线性变换时, 即

$$\Phi(\boldsymbol{x}) = A\boldsymbol{x} + \boldsymbol{b}, \tag{8.3.11}$$

其中 A 为 $m \times m$ 阶方阵, $\boldsymbol{b}$ 为 m 维向量. 于是 $D\Phi(\boldsymbol{x}) = A$, 从而可见, (8.3.10)式右端的极限存在, 且与 $\boldsymbol{x}_0$ 无关. 若

$$0 > \lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_m,$$

即 A 的所有特征值具有小于 1 的模, 则由(8.3.11)式的 Φ 产生的迭代序列必收敛, 且收敛到同一个极限. 在一般情况下, 无论是线性还是非线性变换, Φ 的 Lyapunov 指数 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_m$ 可以有正的, 也有负的, 不必同时为正的, 或者同时为负的.

利用变换 Φ 的 Lyapunov 指数 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_m$ 可以定义 Lyapunov 维数, 即

$$d_L = k - \frac{\lambda_1 + \lambda_2 + \cdots + \lambda_k}{\lambda_{k+1}}, \tag{8.3.12}$$

其中 k 是使 $\lambda_1 + \lambda_2 + \cdots + \lambda_k \geqslant 0$ 的最大正整数; 如果 $\lambda_1 < 0$, 即所有的 $\lambda_i < \lambda_1 < 0$, 则取 $d_L = 0$; 如果 $\lambda_1 + \lambda_2 + \cdots + \lambda_m \geqslant 0$, 则取 $d_L = m$.

例如, 考虑(8.3.11)式的 Φ 的 Lyapunov 指数, 当 $\lambda_1 < 0$ 时, 则 $d_L = 0$; 此时(8.3.11)式中的 A 的特征值的模都小于 1, 根据推论 2.3.10 知, 如下动态系统

$$\begin{cases} \boldsymbol{x}_t = A\boldsymbol{x}_{t-1} + \boldsymbol{b}; \\ \boldsymbol{x}_0 \in R^m, \end{cases} \tag{8.3.13}$$

在 $(I - A)^{-1}\boldsymbol{b}$ 处为几何速度稳定的. 从而系统(8.3.13)的吸引集是一个单点集, 其维数自然是零, 这与 $d_L = 0$ 相一致. 由此例可帮助理解 Lyapunov 维数的含义.

又如, 当(8.3.13)式 φ 的 $A = I, \boldsymbol{b} = \boldsymbol{0}$ 时, 即 $\Phi(\boldsymbol{x}) = \boldsymbol{x}$, 其 Lyapunov 指数为 $\lambda_1 = \lambda_2 = \cdots = \lambda_m = 0$, 从而 $d_L = m$. 注意, 此时 Φ 的吸引集为 R^m, 其维数为 m, 也和 d_L 相一致.

再如, 当(8.3.13)式中的 $m = 2$, 且

$$A = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

易见，此时 $\Phi(\boldsymbol{x})=A\boldsymbol{x}$ 的 Lyapunov 指数为 $\lambda_1=0,\lambda_2=-\log 2$，于是 $d_L=1$. 而此时 Φ 的吸引集为

$$L_\Phi=\{(0,x),\ -\infty<x<\infty\},$$

其维数当然是 1，这又与 d_L 相一致.

在非线性变换情况下，Lyapunov 维数 d_L 不一定总是正整数，也可能是分数.

例如考虑从[0,1]×[0,1]到[0,1]×[0,1]的变换：

$$\Phi(\boldsymbol{x})=\begin{pmatrix}4x_1(1-x_1)\\ \dfrac{1}{4}x_2\end{pmatrix}. \tag{8.3.14}$$

注意 $\boldsymbol{x}=(x_1,x_2)^\tau$，$\Phi(\boldsymbol{x})=(\varphi_1(\boldsymbol{x}),\varphi_2(\boldsymbol{x}))^\tau$，可知

$$\varphi_1(\boldsymbol{x})=\varphi_1(x_1,x_2)=4x_1(1-x_1)\equiv g(x_1), \tag{8.3.15}$$

$$\varphi_2(\boldsymbol{x})=\varphi_2(x_1,x_2)=x_2/4=h(x_2). \tag{8.3.16}$$

由以上三式可知，变换 Φ 是由两个一维变换合并而成的. 于是用(8.3.4)式可分别求出(8.3.16)式中 h 的 Lyapunov 指数 $\lambda_2=\log\dfrac{1}{4}=-2\log 2$ 以及(8.3.15)式中 g 的 Lyapunov 指数 λ_1. 为给出 λ_1 的具体值，先由(8.3.8)和(8.1.10)式，有

$$\lambda_1=\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\log|g'(u_k)|=E\log|g'(u)|,$$

式中 u_k 是指从 u_0 出发，利用(8.3.15)式右端的变换 g 迭代 k 步的取值；而"E"表示按(8.3.15)式中 g 的不变分布(见(4.6.23))Bata(0.5,0.5)求平均运算，于是由(8.3.15)式和 Bata(0.5,0.5)的分布密度，可得到

$$\begin{aligned}\lambda_1=Eg'(u)&=\int_0^1\frac{\log|g'(u)|}{\pi[u(1-u)]^{\frac{1}{2}}}\mathrm{d}u\\&=\int_0^1\frac{\log(4-8u)}{\pi[u(1-u)]^{\frac{1}{2}}}\mathrm{d}u=\log 2.\end{aligned}$$

综上所述，(8.3.14)式中的 Φ 的 Lyapunov 指数为 $\lambda_1=\log 2$，$\lambda_2=-2\log 2$；由(8.3.12)式又知 Φ 的 Lyapunov 维数为 $d_L=\dfrac{3}{2}$，这是一个分数维数.

对于形如(2.2.3)式的 p 阶动态系统，只须考虑相应的一阶高维系统(2.2.4)式，并用上述方法定义其 Lyapunov 指数和维数.

在分形几何学中，除了 Lyapunov 维数外，还有其他分形维数的概

念，现在介绍以下几种.

二、Hausdoff 维数

前面所介绍的 Lyapunov 维数，是借助于变换 Φ 的偏导数和不变分布定义的. 在这一小节中，将直接从集合的特征引入另一种分形维数的定义.

考虑 m 维欧氏空间 R^m 中的有界集合 A，对任给的 $\varepsilon>0$，用直径为 ε 的 R^m 中的球覆盖 A 时，所使用的球的最少个数记为 $N(\varepsilon)$. 如果以下的极限存在，即

$$d_K(A)=\lim_{\varepsilon\to 0}\frac{\log N(\varepsilon)}{\log 1/\varepsilon}, \tag{8.3.17}$$

$d_K(A)$称为集合 A 的 Kolmogorov 容度. 但是，(8.3.17)式右端的极限并不是对任何集合 A 都存在的，为此，用其上极限定义 A 的 Hausdoff 维数，即

$$d_H(A)=\lim_{\varepsilon\to 0}\sup\frac{\log N(\varepsilon)}{\log 1/\varepsilon}. \tag{8.3.18}$$

显然，当(8.3.17)式的极限存在时，$d_K(A)=d_H(A)$.

由于 A 是有界集，故必存在半径为 r 的球，使得

$$S_r=\{\boldsymbol{x}:\boldsymbol{x}\in R^m,\ \|\boldsymbol{x}\|\leqslant r\}\supseteq A.$$

对于 R^m 中的球 S_r，不难验证 $d_H(S_r)=d_K(S_r)=m$. 由于覆盖 S_r 的 $N(\varepsilon)$ 个球也必覆盖其子集合 A，所以必有

$$d_H(A)\leqslant d_H(S_r)=m. \tag{8.3.19}$$

上式表明了 R^m 中的有界集合 A 的 Hausdoff 维数不超过 R^m 的维数 m，此外，根据 Hausdoff 维数的定义，不难看出以下的事实：

如果 A 是 R^m 中的可列点组成的集合，那么必有 $d_K(A)=d_H(A)=0$.

如果 A 是 $R^m(m\geqslant 2)$中的有限长度的直线或曲线，则 $d_K(A)=d_H(A)=1$. 如果 A 是 $R^m(m\geqslant 3)$中的有限面积的平面或曲面，则 $d_K(A)=d_H(A)=2$. 更一般地，如果 A 是 R^m 中的低维超平面或流形，则 $d_H(A)$ 总为小于 m 的整数.

若 A 是[0,1]中的 Cantor 集，可证 $d_K(A)=d_H(A)=\log 2/\log 3$. 详见文献[106].

最后我们指出，对于多维变换 Φ 而言(不一定要求连续可微)，它们

的吸引元、吸引集或吸引域可以用上述方法定义其 Kolmogorov 容度和 Hausdoff 维数等，借以描述这些集合的复杂特征.

三、相关维数

考虑从 R^m 到 R^m 的变换 $\boldsymbol{\Phi}$，以 $\boldsymbol{x}_t$ 表示用 $\boldsymbol{\Phi}$ 迭代 t 步的值，并假定已得到 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 的观测值，计算

$$C_n(r) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} I(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_\infty < r). \tag{8.3.20}$$

式中 $I(\cdot)$ 仍表示示性函数，$0<r<1$；对于向量 $\boldsymbol{y}=(y_1,\cdots,y_m)^\tau$ 和 $\boldsymbol{z}=(z_1,\cdots,z_m)^\tau$，(8.3.20)式中的 $\|\boldsymbol{y}-\boldsymbol{z}\|_\infty$ 被定义为

$$\|\boldsymbol{y}-\boldsymbol{z}\|_\infty = \max\{|y_i - z_i|, i = 1,2,\cdots,m\},$$

当 n 趋于无穷大时，记

$$C(r) = \lim_{n\to\infty} C_n(r). \tag{8.3.21}$$

关于上式中的极限存在问题的讨论与(8.3.8)式的讨论相似，这里从略. 类似于(8.3.17)式，用以下极限定义 $\boldsymbol{\Phi}$ 的相关维数，即

$$d_{GP}(\boldsymbol{\Phi}) = \lim_{r\to 0}\frac{\log C(r)}{\log r}. \tag{8.3.22}$$

关于上述几种不同的分形维数的关系问题，在文献[31]中曾指出

$$d_{GP}(\boldsymbol{\Phi}) \leqslant d_K(L_{\boldsymbol{\Phi}}), \tag{8.3.23}$$

式中 $d_K(L_{\boldsymbol{\Phi}})$ 表示 $\boldsymbol{\Phi}$ 的吸引集的 Kolmogorov 容度. 目前尚不知 d_{GP}、d_K 和 d_L 之间的更多的联系.

四、嵌入维数

本小节要介绍的嵌入维数是与上述的各种分形维数不同的维数概念，但又是有某种联系的. 它是从识别混沌变换的维数角度提出来的.

假如 $\boldsymbol{\Phi}$ 是从 R^m 到 R^m 的变换，$\{\boldsymbol{x}_t, t\geqslant 0\}$ 是满足 $\boldsymbol{x}_t=\boldsymbol{\Phi}(\boldsymbol{x}_{t-1})$ 的迭代序列，由于种种原因此序列不能被直接观测，而代之以观测 $y_t=f(\boldsymbol{x}_t)$，这里 f 是从 R^m 到 R^1 的变换. y_t 可以是 $\boldsymbol{x}_t$ 的某个分量，也可以是 $\boldsymbol{x}_t$ 的某种函数. 我们要问，$\{\boldsymbol{x}_t, t\geqslant 0\}$ 的吸引集的结构能否被 $\{y_t, t\geqslant 0\}$ 再现出来？或者说重新构造出来？现在从反面给这问题一个回答，具体内容如下：一元序列 $\{y_t, t\geqslant 0\}$，它被称为有确定性的和光滑的自释性，如果存在连续可微的变换 $\boldsymbol{\Phi}(R^m\to R^m)$ 和 $f(R^m\to R^1)$，使得 $y_t=f(\boldsymbol{x}_t)$ 且 $\boldsymbol{x}_t=\boldsymbol{\Phi}(\boldsymbol{x}_{t-1})$. 这样的序列具有如下的性质：取 y_t 的 q 个相连的延迟量构成向量序列

$$\boldsymbol{w}_t = (y_t, y_{t-1}, \cdots, y_{t-q+1})^{\tau} \quad (t = q+1, q+2, \cdots),$$

只要

$$q \geqslant 2d_H(L_{\Phi}) + 1, \tag{8.3.24}$$

式中 $d_H(L_{\Phi})$ 为 $\boldsymbol{\Phi}$ 的吸引集的 Hausdoff 维数，则序列 $\{\boldsymbol{w}_t, t\geqslant q\}$ 的极限点集的相关维数与 q 无关. 其实，当(8.3.24)式成立时，$\{\boldsymbol{w}_t, t\geqslant q\}$ 与 $\{\boldsymbol{x}_t, t\geqslant 0\}$ 具有某种相类似的特征(如见文献[31]).

以上事实说明了：如果序列 $\{y_t\}$ 是由某个确定性系统迭代产生的，它必然使(8.3.24)式成立. 但是这并不是充分条件，而且也没有找到能保证 $\{y_t\}$ 为有确定的和光滑的自释性的充分条件.

对于(8.3.24)式还有另一种应用. 假定 d 维向量序列 $\{\boldsymbol{y}_t\}$ 是可观测的，而且 $\boldsymbol{y}_t = T(\boldsymbol{y}_{t-1})$，其中变换 T 的吸引集是一个圆圈，则在此圆圈所在的平面中是无法"看清"此圆圈的全貌的. 好像一只蚂蚁在此平面中看到此圆圈与一只飞鸟所看到的此圆圈的感觉完全不一样(文献[107]). 因为后者是生活在三维空间中. 注意，圆圈的 Hausdoff 维数是 1，以此代入(8.3.24)式右端，可得到不等式 $q\geqslant 3$. 由此可见，至少要生活在包含圆圈在内的 $q(\geqslant 3)$ 维空间中，才能看清圆圈的全貌. 为了观察圆圈的分形几何结构，在 $\boldsymbol{w}_t$ 中取 $q=3$ 也就够了.

将以上直观想法推而广之，令 $\boldsymbol{w}_t = (\boldsymbol{y}_t^{\tau}, \boldsymbol{y}_{t-1}^{\tau}, \cdots, \boldsymbol{y}_{t-q+1}^{\tau})^{\tau}$，以及

$$\boldsymbol{\Phi}(\boldsymbol{w}_{t-1}) = (T^{\tau}(\boldsymbol{y}_{t-1}), \boldsymbol{y}_{t-1}^{\tau}, \cdots, \boldsymbol{y}_{t-q+1}^{\tau})^{\tau}. \tag{8.3.25}$$

若 T 的吸引集为 L_T，它的 Hausdoff 维数为 $d_H(L_T)$，则取 q 为满足

$$q \geqslant 2d_H(L_T) + 1$$

的最小正整数，便可在 R^{qd} 中观察到吸引集 L_T 的全貌；并称此 q 为 L_T 的**嵌入维数**. 详见文献[107].

五、维数的估计问题

前面介绍的几种分形维数，是刻划非线性变换的吸引集的重要量化指标. 例如在第 2 章的例 2.3.25 中，图 2.3.6 所示的闭曲线的长度是无穷的，而它的分形维数是有限的，且是曲线复杂性的一种量化的指标. 当人们得到有限的观测数据序列时，并不知道它们满足怎样的迭代关系，而希望利用这些数据对上述分形维数进行估计，这就是维数估计问题. 解决这样的问题是十分困难的，研究估计的性质也更困难. 下面仅简单介绍相关维数的一种近似估计方法.

记 $x_1, x_2, \cdots, x_n$ 为一维数据序列，不妨假定它们都在 0 和 1 之间，再

令 $r_k=k/S_n$，其中 S_n 是远比 n 小的正整数. 然后，按(8.3.20)式计算 $\log C_n(r_k)$的值，其中 $k=1,2,\cdots,l_n$，l_n 是比 S_n 小的正整数. 用最小二乘法拟合如下的回归模型：

$$\log C_n(r_k) = \alpha\log r_k + \eta_n \quad (k = 1,2,\cdots,l_n).$$

由于 $\log C_n(r_k)$和 r_k 是给定的计算值，于是可得到 α 的最小二乘估计 $\hat{\alpha}_L$，取 $d_{GP}=\hat{\alpha}_L$ 作为相关维数的估计. 至于此估计的性质，则尚待研究.

有关其他估计方法可参见文献[37].

§8.4 时间序列与神经网络

近年来比较活跃的神经网络研究，与统计学有密切的联系，这种联系包括与时间序列分析的联系. 不仅如此，还包括与统计学其他分支的联系，例如与聚类分析、回归分析等分支也有联系. 在这一节中我们着重介绍神经网络与时间序列分析的联系.

神经网络有多种不同类型，以下仅介绍最常用的多层前输网络与时间序列分析的联系. 顾名思义，神经网络是模仿生物神经系统来描述输入输出系统. 以 x 表示输入信息，y 表示输出量，它们之间的联系往往是复杂的和未知的，暂记为 $y=\varphi(x)$. 当 φ 为线性函数时，则属于最简单的一类，此时只要利用(x,y)的若干次观测量$(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)$就可以准确地估计出 φ 来. 这一估计过程在神经网络中称为*学习过程*，又称*训练过程*. 但是，在神经网络中，一般不讨论 φ 为线性的情况. 当 φ 为非线性函数时，上述的学习过程就困难多了，以至很难准确估计出来. 况且在神经网络中的输入输出量多为向量形式，估计 φ 就更困难了. 如在第 7 章所讨论的非线性时间序列的模型估计那样，对 φ 的估计包括多层次的内容，诸如函数类型估计、阶数估计、参数估计等等. 在神经网络分析中，例如多层前输网络，实际上先将 φ 的类型限定，从而使上述的学习过程的内容单一化，即只要估计网络中的未知参数或权函数. 以下将以多输入单输出多层前输网络为例，叙述其定义和学习过程.

考虑 p 元输入 $\boldsymbol{x}$ 和一元输出 y，按三层前输形式的网络联系，如图 8.4.1所示. 在图 8.4.1 中的圆圈记号表示神经元，这些神经元分布在三个不同的层面上. 在图 8.4.1 中共有三层神经元，最左端一层为输入层，右端为输出层，中间为隐蔽层. 神经元之间有信息输送关系，并用带箭头的线段表示；从左向右，前一层的神经元只向下一层输送信息，如图 8.4.

1 中的箭头所示. 不同层中的神经元的个数由设计者选定. 类似地，前输网络的层次也由实际需要决定. 在本节后文中所讨论的内容都是在层次和神经元选定的情况下进行的.

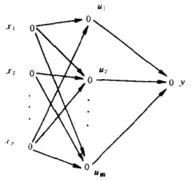


图 8.4.1　三层前输网络

在图 8.4.1 中的三层前输网络中，输入层中共有 p 个神经元，它们将各自的输入量分别输送给下一层中的每个神经元. 在隐蔽层中的每个神经元都从前一层的神经元得到 p 个输入量，即 $x_1,x_2,\cdots,x_p$. 不同的神经元按不同的权数将这 p 个输入量加权求和后作为其纯输入量；然后，又按各自的激发函数产生输出量，并将它输送给下一层的神经元. 最后一层为输出层，此时只有单输出，所以只有一个神经元，它也将前一层输入的诸量加权求和后作为纯输入量，然后再按其激发函数产生输出量 y. 下面用公式描述各层神经元的输入输出关系.

记 $\boldsymbol{x}=(x_1,x_2,\cdots,x_p)^\tau$ 为输入向量，$\boldsymbol{u}=(u_1,u_2,\cdots,u_m)^\tau$ 表示隐蔽层输出向量，即隐蔽层共存 m 个神经元，第 i 个神经元的输出量为 u_i，其纯输入量为

$$\begin{aligned}\psi_i(\boldsymbol{x},\boldsymbol{v}_i) &= v_{i0}+\sum_{j=1}^{p}v_{ij}x_j \\ &= \sum_{j=0}^{p}v_{ij}x_j \quad (i=1,\cdots,m),\end{aligned} \tag{8.4.1}$$

式中 $\boldsymbol{v}_i=(v_{i0},\cdots,v_{ip})^\tau$ 为权向量，但是并不要求其分量 v_{ij} 具有非负性，还约定 $x_0=1$，于是 v_{i0} 表示第 i 个神经元的激发水平调节量. 当 $v_{i1}=v_{i2}=\cdots=v_{ip}=0$ 时，v_{i0} 的调节含义更加突出. 隐蔽层中第 i 个神经元得到纯输入量 $\psi_i(\boldsymbol{x}_i,\boldsymbol{v}_i)$ 后，经过其自身的激发作用给出输出量 u_i，它可以表示为

$$u_i = g_i(\psi_i(\boldsymbol{x}_i,\boldsymbol{v}_i)) \quad (i=1,2,\cdots,m), \tag{8.4.2}$$

其中 g_i 被称为*第 i 个神经元的激发函数*. 类似地，输出层的神经元的纯输入量是前一层的各神经元送来的输入量 $u_1,u_2,\cdots,u_m$ 的加权和，与 (8.4.1) 式相似，即

$$\xi(\boldsymbol{u},\boldsymbol{w}) = w_0+\sum_{j=1}^{m}w_ju_j = \sum_{j=0}^{m}w_ju_j. \tag{8.4.3}$$

其中 $\boldsymbol{w}=(w_0,w_1,\cdots,w_m)^\tau$ 为权向量，并约定 $u_0=1$；由输出神经元的激发

后给出输出量，与(8.4.2)式相似，有

$$y = f(\xi(\boldsymbol{u},\boldsymbol{w})), \tag{8.4.4}$$

其中 f 为输出神经元的激发函数.

综上所述，为了给出三层前输网络系统的输出 y 与输入 $\boldsymbol{x}$ 的联系，还要给出以上诸式 φ 的激发函数 $g_1, g_2, \cdots, g_m$ 和 f，以及所有的权向量 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_m$，和 $\boldsymbol{w}$. 显然，关于激发函数的选择范围太广泛了，在神经网络中，最常用的激发函数有如下两类：一类是符号函数，即

$$g(x) = \begin{cases} 1, & \text{当 } x \geqslant 0; \\ -1, & \text{当 } x < 0. \end{cases} \tag{8.4.5}$$

另一类为 S 形函数：

$$g(x) = 1/(1 + e^{-x}), \ -\infty < x < \infty. \tag{8.4.6}$$

在激发函数选定时，神经网络中的未知权向量便可用学习过程确定，详见后文叙述. 为了说明上述的三层前输网络所能描述的输入输出关系的广泛性，现举以下一些特例说明之.

例 8.4.1 如果 $v_{i0}=0, w_k=1 \ (k=1,2,\cdots,m), f(x)=x$，则(8.4.1)至(8.4.4)式成为

$$\begin{aligned} y &= f(\xi(\boldsymbol{u},\boldsymbol{w})) = w_0 + \sum_{j=1}^{m} u_j \\ &= w_0 + \sum_{i=1}^{m} g_k\Big(\sum_{j=1}^{p} v_{ij} x_j\Big). \end{aligned} \tag{8.4.7}$$

上式模型等价于投影寻踪(Projection Pursuit)模型，参见文献[50].

例 8.4.2 如果 $m=p, \ w_k=1, \ v_{k0}=0, \ v_{kj}=\delta_{kj} \ (1 \leqslant k, \ j \leqslant p)$，其中 $\delta_{kk}=1, \delta_{kj}=0$，当 $k \neq j$. $f(x)=x$. 于是由(8.4.1)～(8.4.4)式，有

$$y = w_0 + \sum_{i=1}^{p} g_i(x_i). \tag{8.4.8}$$

此式为广义可加模型.

例 8.4.3 如果 $g_i = g, i=1,2,\cdots,m$，且 $f(x)=x$，那么由(8.4.1)～(8.4.4)式有

$$y = w_0 + \sum_{i=1}^{m} w_i g(v_{i0} + \sum_{j=1}^{p} v_{ij} x_j). \tag{8.4.9}$$

以上几例中的激发函数 $g_1, \cdots, g_m$ 可取(8.4.5)或(8.4.6)式中的函数，也可取其他非线性函数. 下面再举一例，它不同于以上三例.

例 8.4.4 考虑三层前输网络，其输入输出关系(8.4.1)和(8.4.2)式被以下的形式代替. 首先，隐蔽层中的第 i 个神经元的纯输入量为

$$\| x - c_i \| / \sigma \quad (i = 1,2,\cdots,m), \tag{8.4.10}$$

其中 x 仍如前所述，为输入向量，$c_i=(c_{i1},c_{i2},\cdots,c_{im})^\tau$ 为位置向量，σ 为刻度常数. 另一方面，相应的激发函数为

$$g_i(x) = \sigma^{-p}\varphi(x) \quad (i = 1,2,\cdots,m), \tag{8.4.11}$$

式中 φ 为球对称函数. 于是由第 i 个神经元输给输出神经元的量为 $\sigma^{-p}\varphi(\|x-c_i\|/\sigma)$. 如果输出神经元的纯输入量和输出量仍由(8.4.3)和(8.4.4)式确定，当取 $f(x)=x$ 时，则有

$$y = w_0 + \sum_{j=1}^{m} w_j \sigma^{-p}\varphi(\| x - c_i \| / \sigma). \tag{8.4.12}$$

从函数论观点看，(8.4.12)式所描述的输入 x 与输出 y 的关系，可以作为更一般多输入单输出的逼近. 这种方法被称为*基函数方法*(见文献[21]).

通过以上数例可见，多层前输网络以其特有的方式建立起输入与输出的联系. 特别当所有激发函数只取(8.4.5)或(8.4.6)式的简单函数时，就每个神经元而言，其输入输出的关系是相当简单的非线性关系，而隔层之间的神经元的输入输出关系可以是较复杂的，从而它们覆盖了多种熟知的非线性的输入输出系统. 这正是神经网络描述输入输出系统的重要特征之一.

在神经网络的层次和神经元的个数及激发函数给定后，为了最终确定神经网络的输入输出关系，还要确定(8.4.1)和(8.4.3)式中的权向量. 这可通过前述的学习过程来完成. 换言之，就是通过输入输出的观测数据 (x_t, y_t) $(t=1,2,\cdots,N)$ 对上述诸权向量进行估计，或在原估计基础上进行调整. 这种学习过程与非线性回归分析很相似. 但是，当多层前输网络中的输入量 x 是输出量 y 的时间延迟量时，上述的学习过程则与时间序列分析中的非线性自回归分析相似. 下面介绍带反馈的三层前输网络.

考虑图 8.4.1 所示的三层前输系统为迭代系统，即输出 y 用 y_t 代替，输入 x 用 $x_t=(y_{t-1},y_{t-2},\cdots,y_{t-p})^\tau$ 代替，即将输出 y_t 的延迟量反馈作为输入向量，并重新讨论例 8.4.1～例 8.4.4，于是可得到如下的相应系统：

与(8.4.7)式相应的是

$$y_t = w_0 + \sum_{i=1}^{m} g_i\Big(\sum_{j=1}^{p} v_{ij} y_{t-j}\Big). \tag{8.4.13}$$

与(8.4.8)式相应的是

$$y_t = w_0 + \sum_{i=1}^{p} g_i(y_{t-i}). \tag{8.4.14}$$

与(8.4.9)式相应的是

$$y_t = w_0 + \sum_{i=1}^{m} w_i g\Big(v_{i0} + \sum_{j=1}^{p} v_{ij} y_{t-j}\Big). \tag{8.4.15}$$

与(8.4.12)式相应的是

$$y_t = w_0 + \sum_{i=1}^{m} w_i \sigma^{-p} \varphi\Big(\Big\{\sum_{j=1}^{p}(y_{t-j} - c_{ij})^2\Big\}^{1/2}/\sigma\Big). \tag{8.4.16}$$

显而易见,以上四式是我们所熟悉的非线性动态系统,可与第 2 章的动态系统相对照. 这些系统可以视为如下一般系统的逼近,即

$$y_t = \varphi(y_{t-1}, \cdots, y_{t-p}). \tag{8.4.17}$$

当函数 φ 满足某些光滑性条件时,取适当多的神经元的个数,前面所列的几种系统可对(8.4.17)式达到较好的逼近效果.

为了确定(8.4.13)等式中的权向量,需要安排学习过程,实际上,就是利用观测数据序列 $y_1, y_2, \cdots, y_N$ 估计(8.4.13)等式中的未知权向量. 下面以(8.4.13)式为例,叙述对 w_0, v_{ij} $(1 \leqslant i \leqslant m, 1 \leqslant j \leqslant p)$ 等元素的估计问题. 现记 $\boldsymbol{v}_i = (v_{i1}, \cdots, v_{ip})^\tau$ $(i = 1, 2, \cdots, m)$. 在给定 $w_0, \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_m$ 诸变量时,利用(8.4.13)式和数据 $y_1, y_2, \cdots, y_N$ 可给出拟合值

$$z_t = w_0 + \sum_{i=1}^{m} g_i\Big(\sum_{j=1}^{p} v_{ij} y_{t-j}\Big) \quad (t = p+1, p+2, \cdots, N). \tag{8.4.18}$$

显然,拟合值 z_t 不一定恰好等于(8.4.13)式的真正系统输出量 y_t,除非(8.4.18)式中的诸权数值恰好取真值. 因此,可以定义它们之间的距离 $\rho(y_t, z_t)$,并将它们求和:

$$e(w_0, \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_m) = \sum_{t=p+1}^{N} \rho(y_t, z_t), \tag{8.4.19}$$

其中最常用的距离函数为 $\rho(y_t, z_t) = (y_t - z_t)^2$. 求(8.4.19)式的最小值,其解记为 $\hat{w}_0, \hat{\boldsymbol{v}}_1, \cdots, \hat{\boldsymbol{v}}_m$,它们可作为 $w_0, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_m$ 的估计. 具体求解方法可见神经网络方面的文献.

据前所述,由前输网络所描述的输入输出系统是确定性的系统,但是在用这样的网络描述真实数据资料时,难免带有误差. 从统计观点看,这种误差可视为随机噪声或称随机误差. 以系统(8.4.14)为例,其带误差的形式可写为如下的可加噪声系统:

$$y_t = w_0 + \sum_{i=1}^{p} g_i(y_{t-i}) + \varepsilon_t. \tag{8.4.20}$$

这正是我们在第 4 章所熟悉的可加非线性自回归模型(见(3.2.3)式). 类

似地，其他带反馈的前输神经网络系统的输入输出的关系也有非线性时间序列模型的形式. 这里不一一叙述了.

到此为止，我们不仅看到了神经网络与非线性系统的联系，还看到了与非线性时间序列模型的联系，并把神经网络的学习过程与时间序列建模相联系. 因此，时间序列分析中的建模理论与方法无疑可以被神经网络中的学习过程所引用. 例如本书第 7 章所介绍的某些方法可以作为神经网络的训练方法. 如果需要讨论神经网络的稳定性时，本书第 2 章和第 4 章的内容也有重要参考价值.

第 9 章

实例分析

时间序列分析在科学研究、工业、医学、生物、经济管理以及军事科学领域都有重要应用.线性时间序列分析的方法在这些领域中的应用成果是相当丰富的.但是在实际问题中,有许多现象用线性方法描述是相当粗糙的,往往不能有效地反映和解释所观测的数据的客观规律,这可从本章将要介绍的一些实例中看到.另一方面,实际中的观测数据不一定真的满足一个固定的统计模型.所谓建模分析,只是人们根据统计理论知识和数据的特征,用已知的统计模型和创新的统计模型去拟合这些数据,以图近似地描述观测数据的规律性.因此,对一组数据,我们可以用不同的建模方法去拟合,并从中选择能更好地刻画这些数据特征的模型.本章将通过几个实际数值例子的分析来说明如何使用非线性时间序列分析的方法.

§9.1 加拿大山猫数据的统计分析

加拿大山猫数据,是指从 1821 年～1934 年逐年记录的加拿大马堪泽河(Mackenzie River)流域的捕获山猫的数据.这些数据最早见于文献[44]中.这组数据是时间序列分析中著名的数据.各家新创的时间序列建模方法都以能更精确地预报这组数据作为方法优越性的凭证之一.本节将介绍文献中关于这组数据的一些非线性建模方法.

一、数据的探索性分析

表 9.1.1 列出了 1821～1934 年加拿大捕获山猫数量的资料.由表 9.1.1 中数据的变化可见,加拿大山猫数据有大约 10 年的周期变化,并且每一个局部极小到下一个局部极大的增长周期超过从一个局部极大到下一个局部极小的下降周期.

对原始数据取以 10 为底的对数,把变换后的数据作为建模数据.图 9.1.1 是变换后的数据随时间变化的图形.

表 9.1.1　加拿大山猫数据(1821 年～1934 年)

1-11	12-22	23-33	34-44	45-55	56-66	67-77	78-88	88-99	100-110	111-114
269	98	68	731	3311	1426	389	105	382	108	1000
321	184	213	1638	6721	756	73	153	808	229	1590
585	279	546	2725	4254	299	39	387	1388	399	2657
871	409	1033	2871	687	201	49	758	2713	1132	3396
1475	2285	2129	2119	255	229	59	1307	3800	2432	
2821	2685	2536	684	473	469	188	3465	3091	3574	
3928	3409	957	299	358	736	377	6991	2985	2935	
5943	1824	361	236	784	2042	1292	6313	3790	1537	
4950	409	377	245	1594	2811	4031	3794	674	529	
2577	151	225	552	1676	4431	3495	1836	81	485	
523	45	360	1623	2251	2511	587	345	80	662	

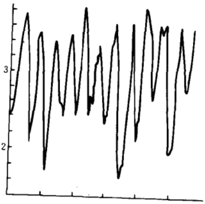


图 9.1.1　加拿大山猫变换数据

变换数据的一种一维直方图由图 9.1.2 给出.它有两个峰值,在点 3 附近有一凹沟,故它不是正态变量的数据.变换数据的样本均值为 2.9036,样本均方差为 0.3090,偏度为 −0.36,峰度为 2.77.

变换数据的几个二维直方图由图 9.1.3 给出,它们呈“山脉”状.

变换数据的二维散点图由图 9.1.4 给出.它显示出一种有规则的变化倾向.延迟步数较小的 (x_t, x_{t-1}) 和 (x_t, x_{t-2}) 的图形(见图9.1.4的(a)和(b))更清晰地表现出数据的周期结构.

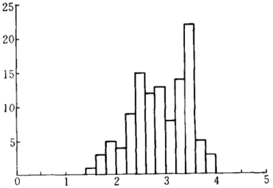


图 9.1.2　变换数据的直方图

图 9.1.3中“最高的山峰”的偏移现象以及图 9.1.4 的(a)和(b)中心的空白现象进一步突出了数据非正态的特征.

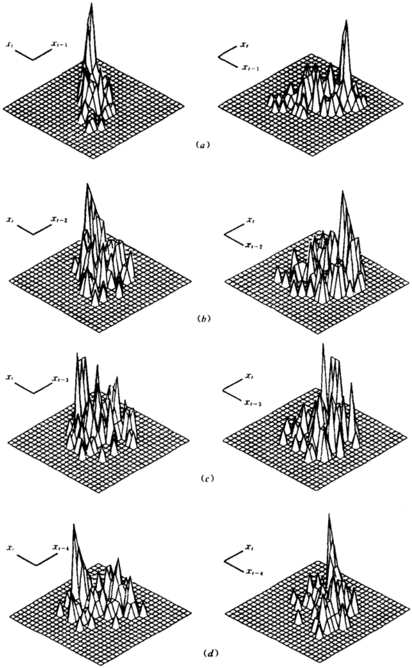


图 9.1.3(1)

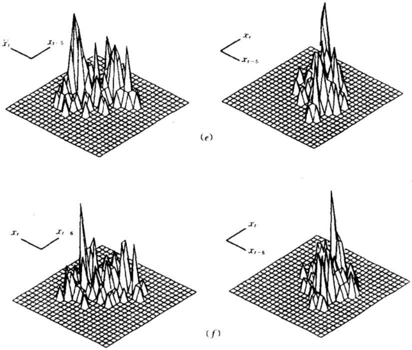


图 9.1.3(2) 变换数据的二维直方图

表 9.1.2 列出了变换数据的样本自相关函数值. 样本自相关的周期变化也反映了数据的周期性. 而且由表中样本自相关的数值可见到其衰减是很慢的.

关于变换数据的线性检验，双谱检验法和 Keenan 检验法在给定置信水平 $\beta=0.05$ 时，不能拒绝变换数据的线性性假定. 而 § 6.2 中的 K-S 检验在给定置信水平 $\beta=0.05$ 时，拒绝变换数据的线性性假定.

表 9.1.2 变换数据的样本自相关

j	$\hat{\rho}_j$	j	$\hat{\rho}_j$	j	$\hat{\rho}_j$
1	0.78513	10	0.60551	19	0.45500
2	0.34024	11	0.38295	20	0.44650
3	−0.13229	12	−0.01229	21	0.21760
4	−0.49389	13	−0.38480	22	−0.11551
5	−0.62055	14	−0.60734	23	−0.39955
6	−0.48796	15	−0.61020	24	−0.50749
7	−0.15781	16	−0.40691	25	−0.41462
8	0.23485	17	−0.07272	26	0.16318
9	9.53721	18	0.25318	27	0.15091

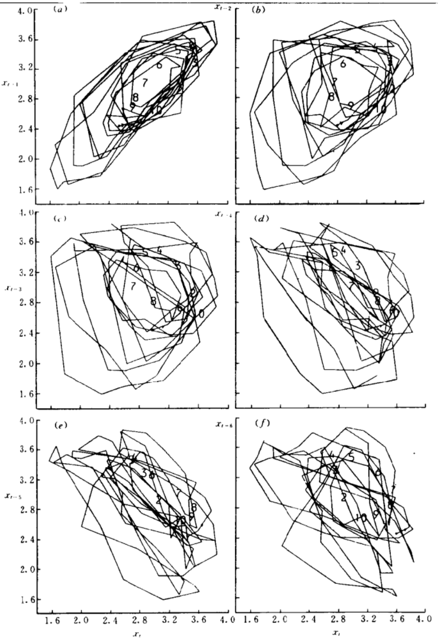


图 9.1.4　变换数据的二维散点图

二、变换数据的统计建模

Moran(见文献[80])首先对山猫数据进行了统计分析,提出了如下的 LAR(2)模型:

$$x_t - 2.9036 = 1.4101(x_{t-1} - 2.9036) - 0.7734(x_{t-2} - 2.9036) + \varepsilon_t, \qquad (9.1.1)$$

式中 ε_t 的方差估计为 $\hat{\sigma}_\varepsilon^2=0.0459$，LAR(2)模型(9.1.1)的系数多项式为

$$\varphi(z)=1-1.4101z+0.7734z^2.$$

$\varphi(z)$的零点为$(0.8794)^{-1}\exp\{\pm 2\pi i/9.8076\}$. 因此，满足(9.1.1)的LAR(2)过程能显示出具有周期9.8076的周期循环. 但是模型(9.1.1)的自相关函数的衰减比变换数据的自相关函数要快得多. 这就表明了对刻画山猫数据的变化规律而言，模型(9.1.1)不是一个好模型. Moran(1953)曾指出："也许……通过某种非线性模型，过程(指山猫数据，作者注)将会被很好地描述".

Tong(见文献[103])考虑用一个高阶疏系数LAR模型拟合山猫数据. 用AIC准则得到一个LAR(12)模型，并得到如下12阶疏系数LAR模型：

$$\begin{aligned} x_t' = {} & 1.0569x_{t-1}' - 0.3374x_{t-2}' - 0.0945x_{t-4}' \\ & + 0.1512x_{t-9}' - 0.1759x_{t-12}' + \varepsilon_t, \end{aligned} \tag{9.1.2}$$

其中 $\hat{\sigma}_\varepsilon^2=0.0503$，$x_t'=x_t-2.8802$. 尽管模型(9.1.2)对数据拟合得较好，但是有些参数很难从生物的观点给予解释. 特别是，一只山猫存活10年，被认为是很稀罕的，而且不论山猫出生在哪一年，其年死亡率为60%.

Subba Rao 和 Gabr(见文献[99])也用高阶LAR模型拟合山猫数据，得到如下的LAR(12)模型：

$$\begin{aligned} x_t = {} & 1.0541x_{t-1} - 0.4539x_{t-2} + 0.3260x_{t-3} \\ & - 0.3791x_{t-4} + 0.2345x_{t-5} - 0.1757x_{t-6} \\ & + 0.096x_{t-7} - 0.1284x_{t-8} + 0.2744x_{t-9} \\ & + 0.1143x_{t-10} - 0.1853x_{t-11} - 0.1713x_{t-12} + \varepsilon_t. \end{aligned} \tag{9.1.3}$$

和12阶的疏系数LAR模型：

$$\begin{aligned} x_t = {} & 1.0170x_{t-1} - 0.4900x_{t-2} + 0.2585x_{t-3} \\ & - 0.2204x_{t-4} + 0.2210x_{t-9} - 0.2534x_{t-12} + \varepsilon_t. \end{aligned} \tag{9.1.4}$$

Tong 和 Lim(见文献[108])建立了如下的门限自回归模型：

$$x_t=\begin{cases} 0.5239+1.0359x_{t-1}-0.1756x_{t-2} \\ +0.1753x_{t-3}-0.4339x_{t-4}+0.3457x_{t-5} \\ -0.3032x_{t-6}+0.2165x_{t-7}+0.0043x_{t-8}+\varepsilon_{1t}, \\ \qquad \text{当 } x_{t-2}\leqslant 3.1163; \\ 2.6559+1.4246x_{t-1}-1.1618x_{t-2}-0.1094x_{t-3}+\varepsilon_{2t}, \\ \qquad \text{当 } x_{t-2}>3.1163, \end{cases} \tag{9.1.5}$$

其中 $E\varepsilon_{1t}^2=0.0255$，$E\varepsilon_{2t}^2=0.0516$.

模型(9.1.5)是用 1821 年～1920 年的 100 个数据求条件最小二乘估计所得到. 利用模型(9.1.5)，对 1921 年～1934 年山猫捕获量进行预报，并与实际数值进行比较(见表 9.1.3). 图 9.1.5 绘出了用(9.1.5)式进行预报的结果. 由表 9.1.3 可知，模型(9.1.5)给出的一步预报对实测数据具有较好的拟合度，而且当模型进入稳定状态后，一步预报值构成的最终预报函数是一个以 9 年为周期的非对称周期函数，其中上升周期为 6，下降周期为 3，从而构成周期为 9 的极限环. 极限环为(2.6226，2.8945，3.2523，3.4601，3.4257，3.2281，2.9793，2.7884，2.6639)(见图 9.1.5). 这些预报值按(x_t, x_{t-1})的散点图(见图 9.1.6)的轨迹与变换数据的散点图的轨迹(见图 9.1.4(a))很相似.

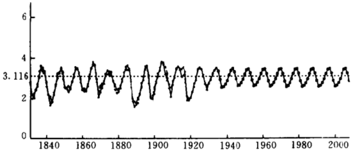


图 9.1.5　山猫捕获量预报

由模型(9.1.5)构成的极限环成功地解释了山猫在自然界中生态平衡的规律. 周期为 9 的极限环的出现与生态学中普遍认为的山猫生态 9～10 年周期的结论相一致. $d=2$说明当年的群体数目主要与两年前的群体数目有特殊联系，而山猫从出生到具有生殖能力的成熟期正好是两年. 可见$d=2$有明显的生态意义. 进一步将模型(9.1.5)近似地简化为

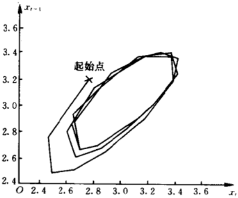


图 9.1.6　对山猫数据所建 TAR 模型的极限环

$$\nabla x_t = x_t - x_{t-1}$$

$$= \begin{cases} 0.5, & \text{当 } x_{t-2} \leqslant 3; \\ 2.6 + 0.4x_{t-1} - x_{t-2}, & \text{当 } x_{t-2} > 3. \end{cases} \qquad (9.1.6)$$

(9.1.6)式形象地刻划了山猫增减与所需食品之间的相互制约的关系.当某年山猫数量(经对数变换后)不多于 3 时,则意味着所需食品大于山猫的生存需求,从而导致两年后成熟的山猫繁殖能力旺盛,山猫的总数趋于增加.当其超过 3 时,所需食品相对地缺乏,造成成年山猫和幼小山猫因饥饿而死,以及饥饿的成年山猫生殖能力降低,使得山猫的总数日趋减少.当其数量再降到门限值 3 之内时,又因食品相对丰富,再次导致山猫数量增加,形成上述的循环,如此周而复始.

表 9.1.3 山猫数据预报结果比较

年	实测数据	LAR(12)	Moran	TAR	BL
1921	2.3598	2.4550	2.4504	2.3109	2.4420
1922	2.6010	2.8070	2.8099	2.8770	2.7560
1923	3.0539	2.8990	2.8974	2.9106	2.8970
1924	3.3860	3.2310	3.3495	3.3703	3.1350
1925	3.5532	3.3880	3.4676	3.5875	3.4110
1926	3.4676	3.3320	3.4465	3.4261	3.5120
1927	3.1867	3.0070	2.1966	3.0936	2.9220
1928	2.7235	2.6880	2.8666	2.7706	2.7060
1929	2.6857	2.4280	2.4307	2.4217	2.5830
1930	2.8209	2.7650	2.7357	2.7644	2.8440
1931	3.0000	2.9840	2.9554	2.9397	2.9660
1932	3.2014	3.2170	3.1036	3.2462	3.1590
1933	3.4244	3.3650	3.2490	3.3701	3.2990
1934	3.5310	3.5030	3.4077	3.4468	3.4150

注:LAR(12) 指模型(9.1.2)
Moran 指模型(9.1.1)
TAR 指模型(9.1.5)
BL 指模型(9.1.7)

对山猫数据的非线性拟合还有 Gabr 和 Subba Rao(见文献[51])的双线性模型方法,他们得到如下模型:

$$\begin{aligned} x_t &+ \hat{a}_1 x_{t-1} + \hat{a}_2 x_{t-2} + \hat{a}_3 x_{t-3} + \hat{a}_4 x_{t-4} + \hat{a}_9 x_{t-9} + \hat{a}_{12} x_{t-12} \\ &= \hat{\alpha} + \hat{b}_{39} x_{t-3}\varepsilon_{t-9} + \hat{b}_{99} x_{t-9}\varepsilon_{t-9} + \hat{b}_{62} x_{t-6}\varepsilon_{t-2} \\ &\quad + \hat{b}_{11} x_{t-1}\varepsilon_{t-1} + \hat{b}_{27} x_{t-2}\varepsilon_{t-7} + \hat{b}_{42} x_{t-4}\varepsilon_{t-2} + \varepsilon_t, \end{aligned} \qquad (9.1.7)$$

其中

$\hat{a}_1=-0.77277$， $\hat{a}_2=0.09157$， $\hat{a}_3=-0.83070$，
$\hat{a}_4=0.26149$， $\hat{a}_9=-0.22559$， $\hat{a}_{12}=0.24584$，
$\hat{\alpha}=-1.48629$， $\hat{b}_{39}=-0.78930$， $\hat{b}_{99}=0.47980$，
$\hat{b}_{62}=0.39020$， $\hat{b}_{11}=0.13260$， $\hat{b}_{27}=0.07944$，
$\hat{b}_{42}=-0.32120$， $\hat{\sigma}_\varepsilon^2=0.0223$.

模型(9.1.7)是用1821年～1920年的山猫变换数据拟合的.其余的14个观测值留作与模型(9.1.7)输出的预报值相比较(见表9.1.3).

§9.2 太阳黑子数据的统计分析

太阳黑子的活动直接影响着地球气候的变化.太阳黑子聚集到一定程度就会产生耀斑,释放出的大量能量所产生的电磁辐射和粒子辐射传播到地球,可造成无线电通讯突然中断.因此,准确地预报未来时刻的太阳黑子数是很有意义的,从而也是新创时间序列建模方法都要为之一试的.表9.2.1列出了1700年～1988年历年的太阳黑子平均数,其随时间变化的图形见图9.2.1.

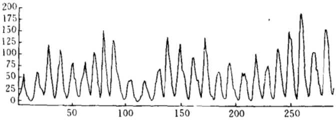


图9.2.1 太阳黑子数据(1700年～1987年)

历史上最早用时间序列研究太阳黑子数据的学者是Yule.他于1927年提出了LAR模型,并用这一模型拟合太阳黑子数据,借以描述太阳黑子的变化规律.50年代以来,这方面的有关研究文献激增.我国的统计学者和天文工作者在这方面也做了许多工作.关于太阳黑子的线性序列建模方法见文献[128],这里从略 本节主要介绍太阳黑子数据的非线性时间序列建模方法.

Tong和Lim(见文献[108])用门限自回归模型拟合了1700年～1920年间的太阳黑子数据,其模型如下:

表 9.2.1　1700 年～1988 年逐年太阳黑子数据

1-28	29-56	57-84	85-112	113-140	141-168	169-196	197-224	225-252	253-280	281-288
5.0	103.0	10.2	10.2	5.0	64.6	37.6	41.8	16.7	31.5	154.7
11.0	73.0	32.4	24.1	12.2	36.7	74.0	26.2	44.3	13.9	140.5
16.0	47.0	47.6	82.9	13.9	24.2	139.0	26.7	63.9	4.4	115.9
23.0	35.0	54.0	132.0	35.4	10.7	111.2	12.1	69.0	38.0	66.6
36.0	11.0	62.9	130.9	45.8	15.0	101.6	9.5	77.8	141.7	45.9
58.0	5.0	85.9	118.1	41.1	40.1	66.2	2.7	64.9	190.2	17.9
29.0	16.0	61.2	89.9	30.1	61.5	44.7	5.0	35.7	184.8	13.4
20.0	34.0	45.1	66.6	23.9	98.5	17.0	24.4	21.2	159.0	29.2
10.0	70.0	36.4	60.0	15.6	124.7	11.3	42.0	11.1	112.3	100.2
8.0	81.0	20.9	46.9	6.6	96.3	12.4	63.5	5.7	53.9	
3.0	111.0	11.4	41.0	4.0	66.6	3.4	53.8	8.7	37.5	
0.0	101.0	37.8	21.3	1.8	64.5	6.0	62.0	36.1	27.9	
0.0	73.0	69.8	16.0	8.5	54.1	32.3	48.5	79.7	10.2	
2.0	40.0	106.1	6.4	16.6	39.0	54.3	43.9	114.4	15.1	
11.0	20.0	100.8	4.1	36.3	20.6	59.7	18.6	109.6	47.0	
27.0	16.0	81.6	6.8	49.6	6.7	63.7	5.7	88.8	93.8	
47.0	5.0	66.5	14.5	64.2	4.3	63.5	3.6	67.8	105.9	
63.0	11.0	34.8	34.0	67.0	22.7	52.2	1.4	47.5	105.5	
60.0	22.0	30.6	45.0	70.9	54.8	25.4	9.6	30.6	104.5	
39.0	40.0	7.0	43.1	47.8	93.8	13.1	47.4	16.3	66.6	
28.0	60.0	19.8	47.5	27.5	95.8	6.8	57.1	9.6	68.9	
26.0	80.9	92.5	42.2	8.5	77.2	6.3	103.9	33.2	38.0	
22.0	83.4	154.4	28.1	13.2	59.1	7.1	80.6	92.6	34.5	
11.0	47.7	125.9	10.1	56.9	44.0	35.6	63.6	151.6	15.5	
21.0	47.8	84.8	8.1	121.5	47.0	73.0	37.6	136.3	12.6	
40.0	30.7	68.1	2.5	138.3	30.5	85.1	26.1	134.7	27.5	
78.0	12.2	38.5	0.0	103.2	16.3	78.0	14.2	83.9	92.5	
122.0	9.6	22.8	1.4	85.7	7.3	64.0	5.8	69.4	155.4	

$$
x_t = \begin{cases}
10.5440 + 1.6920x_{t-1} - 1.1592x_{t-2} \\
+ 0.2367x_{t-3} + 0.1503x_{t-4} + \varepsilon_{1t}, \\
\qquad 当\ x_{t-3} \leqslant 36.6; \\
7.8041 + 0.7432x_{t-1} - 0.0409x_{t-2} \\
- 0.2020x_{t-3} + 0.1730x_{t-4} - 0.2236x_{t-5} \\
+ 0.0189x_{t-6} + 0.1612x_{t-7} - 0.2564x_{t-8} \\
+ 0.3195x_{t-9} - 0.3891x_{t-10} + 0.4306x_{t-11} \\
- 0.0397x_{t-12} + \varepsilon_{2t}, \\
\qquad 当\ x_{t-3} > 36.6.
\end{cases}
\tag{9.2.1}
$$

其中 $\hat{\sigma}_1^2 = 254.64$，$\hat{\sigma}_2^2 = 66.84$.

用模型(9.2.1)对 1921 年～1955 年的太阳黑子数进行预报. 经实际计算结果表明，实测值、用模型(9.2.1)计算出的拟合值以及一步预报值，三者很接近，而且 1921 年以后的预报值逐渐呈现出有规则的循环，并显示出有长度为 31 年的周期，每个周期内又由三个极大和三个极小构成三个局部循环，且局部循环的上升和下降的长度是不对称的. 这又进一步揭示了太阳黑子数目的分周期应为 31/3≈10.3(年).

Tong(1980)对 1700 年～1979 年太阳黑子数据用门限自回归模型建模，并对 1980 年～1987 年太阳黑子数作预报. 先对太阳黑子数据进行 Box-Cox 变换，即对原始数据 $x_t(t=1,\cdots,280)$令

$$z_t(\lambda) = \begin{cases} \dfrac{x_t^\lambda - 1}{\lambda}, & \lambda \neq 0; \\ \log x_t, & \lambda = 0, \end{cases}$$

式中$|\lambda| \leqslant 1$(λ 称为 Box-Cox 参数)，可取为 $\lambda = 0, 0.1, 0.2, \cdots, 1$，分别计算出新序列$\{z_t(\lambda)\}$，并对其建模，求出相应的 $\mathrm{AIC}^{(\lambda)}$值. 令

$$\mathrm{AIC}(\lambda) = \mathrm{AIC}^{(\lambda)} - 2\log|J(\lambda)|,$$

式中

$$J(\lambda) = \prod_{t=1}^{280} x_t^{\lambda-1}.$$

以满足下式的 $\hat{\lambda}$ 作为变换参数 λ 的估计：

$$\mathrm{AIC}(\hat{\lambda}) = \min_{0 \leqslant \lambda \leqslant 1} \mathrm{AIC}(\lambda).$$

对 1700 年～1979 年的太阳黑子数据，经 Box-Cox 变换选 $\hat{\lambda} = \dfrac{1}{2}$，此时

$$z_t = 2(\sqrt{x_t} - 1) \quad (t = 1, \cdots, 280).$$

对$\{z_t\}$建门限自回归模型，得到

$$z_t = \begin{cases} 1.92 + 0.84z_{t-1} + 0.07z_{t-2} + 0.32z_{t-3} \\ + 0.15z_{t-4} - 0.20z_{t-5} - 0.001z_{t-6} + 0.19z_{t-7} \\ - 0.27z_{t-8} + 0.21z_{t-9} + 0.01z_{t-10} + 0.09z_{t-11} + \varepsilon_{1t}, \\ \qquad \text{当 } z_{t-8} \leqslant 11.93; \\ 4.27 + 1.44z_{t-1} - 0.84z_{t-2} + 0.06z_{t-3} + \varepsilon_{2t}, \\ \qquad \text{当 } z_{t-8} > 11.93. \end{cases} \tag{9.2.2}$$

由表 9.2.2 列出的数值可见，建模的效果是理想的，即模型(9.2.2)

能较好地刻画太阳黑子的变化规律.

表 9.2.2　太阳黑子数预报值与实测值比较(1700 年～1979 年数据建模)

年	预报值	实测值	误差
1980	160.1	154.7	−5.4
1981	141.8	140.5	−1.3
1982	96.4	115.9	19.5
1983	61.8	66.6	4.8
1984	31.1	45.9	14.8
1985	18.1	17.9	−0.2
1986	18.9	13.4	−5.5
1987	29.9	29.2	−0.7

Subba Rao 和 Gabr(见文献[99])用疏系数双线性模型拟合太阳黑子数据,得到如下模型:

$$\begin{aligned} x_t + \hat{a}_1 x_{t-1} + \hat{a}_2 x_{t-2} + \hat{a}_9 x_{t-9} &= \hat{\alpha} + \hat{b}_{21} x_{t-2}\varepsilon_{t-1} + \hat{b}_{81} x_{t-8}\varepsilon_{t-1} + \hat{b}_{13} x_{t-1}\varepsilon_{t-3} + \hat{b}_{43} x_{t-4}\varepsilon_{t-3} \\ &\quad + \hat{b}_{16} x_{t-1}\varepsilon_{t-6} + \hat{b}_{24} x_{t-2}\varepsilon_{t-4} + \hat{b}_{32} x_{t-3}\varepsilon_{t-2} + \varepsilon_t, \end{aligned} \tag{9.2.3}$$

式中

$$\begin{aligned} &\hat{a}_1 = -1.5012, \quad \hat{a}_2 = 0.7670, \quad \hat{a}_3 = -0.1152, \\ &\hat{\alpha} = 6.886, \quad \hat{b}_{21} = -0.1458, \quad \hat{b}_{81} = 0.0063, \\ &\hat{b}_{13} = -0.0072, \quad \hat{b}_{43} = -0.0060, \quad \hat{b}_{16} = 0.0036, \\ &\hat{b}_{24} = 0.0043, \quad \hat{b}_{32} = 0.0018. \end{aligned}$$

并用模型(9.2.3)进行了一步预报,有关结果参见文献[99].

我国的天文工作者也运用门限自回归模型拟合太阳黑子数据,并用以预报太阳黑子数的长期变化.这方面的材料有文献[52]、[130]等.以下简要介绍文献[130]中张桂清的工作.

对太阳黑子数据作如下变换:

$$z_t = (x_t + 1)^{1/2}.$$

数据起始点为 1700 年,从 1956 年开始,每增加两年数据,便拟合新的门限自回归模型.所建模型和对 1956 年～1985 年太阳黑子数作预报的结果,分别列在表 9.2.3 和表 9.2.4 中.有趣的是:由表 9.2.3 可见延迟参数 d 呈周期变化,周期为 10 年;而且太阳活动峰年期间延迟参数值较大,例如 1968 年和 1979 年为峰年,相应的延迟参数分别为 8 和 5,表 9.2.4 列出了 1956 年～1985 年共 30 年的预报结果.经与三个太阳活动周的实

测值比较，预报效果是理想的. 按天文学 1956 年～1985 年为太阳活动的第 19、20、21 周，三个太阳活动周实测的极大年分别为 1957 年、1968 年和 1979 年，极小年分别是 1964 年和 1976 年. 预报所得结果是：极大年分别为 1958 年、1968 年和 1980 年；极小年分别为 1965 年和 1977 年. 极大年的预测值中有两个滞后一年，极小年的预测值都滞后一年. 但总的效果仍比其他预报方法好.

表 9.2.3　1956 年～1985 年的模型参数

时间(年)	d	r	k_1	k_2	$\theta^{(1)}=(a_0^{(1)}a_1^{(1)}\cdots a_p^{(1)})^\tau$	$\theta^{(2)}=(a_0^{(2)}a_1^{(2)}\cdots a_p^{(2)})^\tau$
1956 1957	5	6.928	14	12	(1.4180　1.1310　−0.2888 −0.2975　0.3278　−0.2752 0.0002　0.0258　0.1340 0.0827　0.0820　0.1382 −0.0790　−0.1266　0.0843)	(1.9930　1.3560　−0.6973 −0.2784　0.3171　−0.3876 0.0274　0.2604　0.4664 0.6463　−0.2723　−0.0028 0.2926)
1958 1959	8	4.243	11	9	(1.6100　0.9153　−0.2372 −0.1968　0.3668　−0.4284 0.0541　0.1365　0.2923 0.1579　−0.0569　0.2054)	(0.3327　1.2630　−0.5112 −0.1374　0.1958　−0.1269 −0.1259　0.1676　0.0417 0.1537)
1960 1961	4	6.964	8	11	(0.7861　1.3180　−0.5683 −0.0865　0.2031　−0.1586 0.1690　0.2271　0.1145)	(1.6570　0.8418　−0.1824 −0.2628　0.3773　−0.3058 −0.0223　0.1540　−0.2716 0.3817　−0.2278　0.2047)
1962 1963	4	6.285	8	10	(0.4298　1.3110　−0.5624 −0.0943　0.2666　0.0141 −0.3209　0.1691　0.2030)	(1.1370　−0.8281　−0.1548 −0.1258　0.3503　−0.3427 0.0199　0.1983　−0.3783 0.4635　−0.1286)
1964 1965	4	6.285	10	9	(0.3765　1.2940　−0.5547 −0.0879　0.2441　0.0895 −0.2724　0.1381　0.1756 −0.0831　0.0852)	(1.0990　0.7895　−0.1345 −0.1351　0.3434　−0.3293 −0.0018　0.2271　−0.3377 0.3193)
1966 1967	5	6.285	10	10	(1.2570　1.0180　−0.1800 −0.3129　0.2455　−0.1777 0.0002　0.0011　0.2259 −0.0922　0.1018)	(0.3952　1.4540　−0.7396 −0.0550　0.4279　−0.4440 0.0961　0.3368　−0.5731 0.7099　−0.2362)
1968 1969	8	6.964	10	2	(1.2650　0.8686　0.0263 −0.2812　0.1571　−0.2184 0.0017　0.1928　−0.2834 0.1961　0.1007)	(2.6670　1.3970　−0.7706)

（续表）

时间（年）	d	r	k_1	k_2	$\theta^{(1)}=(a_0^{(1)}a_1^{(1)}\cdots a_{p_1}^{(1)})^{\tau}$	$\theta^{(2)}=(a_0^{(2)}a_1^{(2)}\cdots a_{p_2}^{(2)})^{\tau}$
1970 1971	4	6.403	8	9	(0.7035　1.2840　−0.5831 −0.0391　0.2254　−0.0442 −0.2695　0.2111　0.1583)	(1.0980　0.7793　−0.1153 −0.1189　0.3147　−0.3252 0.0088　0.2141　−0.3229 0.3027)
1972 1973	4	6.986	7	13	(0.9436　1.3330　−0.6426 0.0194　0.1260　−0.1296 −0.2152　0.3781)	(0.7408　0.5881　0.0875 −0.2277　0.3191　−0.3546 −0.0007　0.1847　−0.2535 0.3017　−0.0939　−0.0253 0.1036　0.1230)
1974 1975	4	7.036	7	12	(0.9008　1.3370　−0.6398 0.0094　0.1373　−0.1434 −0.1990　0.3729)	(0.8268　0.5735　0.1496 −0.2690　0.2883　−0.2961 −0.0218　0.1895　−0.2554 0.3015　−0.0905　0.0801 0.2654)
1976 1977	4	6.284	8	9	(0.4705　1.3110　−0.5703 −0.0768　0.2615　−0.0229 −0.2851　0.1878　0.1742)	(1.0260　0.7546　−0.0562 −0.1629　0.3178　−0.3145 −0.0083　0.2583　−0.3772 0.3500)
1978 1979	5	8.031	15	12	(1.3800　1.1760　−0.3842 −0.2049　0.3077　−0.3626 0.1855　−0.0595　0.0067 0.0946　0.0230　0.1125 0.0050　−0.2856　0.2928 −0.1334)	(2.8850　1.2520　−0.8703 −0.3333　0.5282　−0.4190 −0.0623　0.3679　−0.6659 0.7548　−0.2218　−0.0296 0.3524)
1980 1981	8	6.411	11	9	(1.3240　0.7930　−0.0157 0.2317　0.1948　−0.2514 −0.0039　0.2069　−0.3321 0.2288　−0.0141　0.1489)	(1.0040　1.3460　−0.7062 −0.0174　0.2560　−0.2531 −0.1418　0.2429　−0.0315 0.1389)
1982 1983	4	6.403	8	9	(0.6545　1.2930　−0.5927 −0.0491　0.2596　−0.0758 −0.2243　0.1691　0.1767)	(1.0560　0.7574　−0.0457 −0.1584　0.3033　−0.2939 −0.0304　0.2529　−0.3452 0.3048)
1984 1985	4	7.409	7	12	(0.8106　1.3500　−0.6534 −0.0060　0.1990　−0.2417 0.1501　0.3849)	(0.8042　0.5447　0.1958 −0.2824　0.2910　−0.2980 −0.0352　0.1919　0.2085 0.2151　−0.0497　−0.0888 0.2723)

表 9.2.4 1956 年～1985 年太阳黑子相对数年均值及对它的预报

时间(年)	(R)实测	($\hat{R}$)预报	绝对误差	时间(年)	(R)实测	($\hat{R}$)预报	绝对误差	平均误差(ΔR)*)
1956	141.7	107.7	34.0	1971	66.6	59.0	7.6	
1957	190.2	158.3	31.9	1972	68.9	53.3	15.6	
1958	184.8	165.8	19.0	1973	38.0	40.3	−2.3	
1959	159.0	151.0	8.0	1974	34.5	27.1	7.4	
1960	112.3	117.1	−4.8	1975	15.5	23.7	−8.2	
1961	53.9	78.1	−24.2	1976	12.6	22.6	−10.0	
1962	37.5	31.7	5.8	1977	27.5	21.1	6.4	
1963	27.9	25.5	2.4	1978	92.5	77.3	15.2	
1964	10.2	15.6	−5.4	1979	155.4	114.8	40.6	
1965	15.1	15.4	−0.3	1980	154.6	152.4	2.2	
1966	47.0	48.0	−1.0	1981	140.4	126.0	14.4	
1967	93.8	72.0	21.8	1982	115.9	91.9	24.0	
1968	105.9	118.4	−12.5	1983	66.6	66.3	0.3	
1969	105.5	107.7	−2.2	1984	45.9	36.9	9.0	
1970	104.5	81.5	23.0	1985	17.9	33.0	15.1	12.5

*) $\Delta R = \left(\sum_{i=1}^{20} |R_i - \hat{R}_i|\right) \Big/ 30$

§9.3 其他实例分析

本节介绍使用非线性时间序列模型的另外两个例子曾在国内有关刊物刊载.

一、南方涛动指数的非线性建模

年际尺度的气候变化中,最突出的形式是南方涛动. 它是赤道地面气压的大尺度位移的标志. 以印度尼西亚、澳大利亚和太平洋地区最为显著. 南方涛动指数(SOI)就是在气压差的基础上确定的. 当东太平洋气压高,而印尼-澳大利亚气压相对低于长期平均值,则称高指数;反之,称为低指数. 由于南方涛动与赤道太平洋海表温度变化有明显的耦合,在海-气相互作用的研究中,SOI 是很受重视的时间序列.

将 1935 年 1 月～1986 年 8 月的 SOI 资料进行 5 个月滑动平均,得到 $n=616$ 个数据. 对这组数据分别拟合 LAR 模型、叠加模型和门限自回归模型,分别得到如下模型:

LAR(12)模型:

$$\begin{aligned} x_t = & 1.2981x_{t-1} - 0.1993x_{t-2} - 0.060x_{t-3} \\ & + 0.0018x_{t-4} - 0.6383x_{t-5} + 0.7881x_{t-6} \\ & - 0.2250x_{t-7} - 0.0061x_{t-8} - 0.0173x_{t-9} \\ & - 0.3433x_{t-10} + 0.3929x_{t-11} - 0.1053x_{t-12} + \varepsilon_t, \end{aligned} \tag{9.3.1}$$

经计算,拟合值与实测值的相关系数 $\rho=0.102$,可见 LAR(12)模型(9.3.1)对 SOI 数据的拟合效果很差.

叠加模型:

考虑到 SOI 资料的趋势性,先用方差分析提取 5～7 年左右的三个周期,将分离周期趋势后剩余的新序列$\{x_t'\}$建立 LAR(3)模型,即

$$\begin{cases} x_t = f(t) + x_t'; \\ x_t' = 0.2458x_{t-1}' + 0.1977x_{t-2}' + 0.1544x_{t-3}' + \varepsilon_t. \end{cases} \tag{9.3.2}$$

模型(9.3.2)的拟合值与实测值的相关系数为 $\rho=0.582$. 可见,对 SOI 资料,模型(9.3.2)比模型(9.3.1)的拟合效果好.

门限自回归模型:

$$x_t = \begin{cases} -0.2624 + 0.9613x_{t-1} - 0.3612x_{t-2} \\ \quad - 0.1460x_{t-3} - 0.2068x_{t-4} - 0.0861x_{t-5} \\ \quad + 0.2677x_{t-6} + \varepsilon_1 t, \quad 当\ x_{t-4} \leqslant 0.10; \\ -0.1618 + 0.9651x_{t-1} - 0.4250x_{t-2} \\ \quad - 0.2793x_{t-3} - 0.1141x_{t-4} - 0.1351x_{t-5} \\ \quad + 0.3245x_{t-6} + \varepsilon_{2t}, \quad 当\ x_{t-4} > 0.10. \end{cases} \tag{9.3.3}$$

模型(9.3.3)的拟合值与实测值之间的相关系数为 $\rho=0.925$. 显然,门限自回归模型的拟合能力比上述两个线性模型有显著提高.

二、ARCH 模型在中国宏观经济分析中的应用

在传统的计量经济学方法中,对宽平稳 LAR(p)模型

$$x_t = \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + e_t \tag{9.3.4}$$

常假定:

(i) 系数 $\alpha_1,\cdots,\alpha_p$ 满足(1.2.5)式;

(ii) $\{e_t\}$为平稳鞅差序列,满足:

$$E\{e_t|\mathscr{F}_{t-1}^x\} = 0,\text{a.s.},\ E\{e_t^2|\mathscr{F}_{t-1}^x\} = \sigma^2,\ \text{a.s.}. \tag{9.3.5}$$

实际上,若记

$$e_t = x_t - E\{x_t | \mathscr{F}_{t-1}^x\},$$

则(9.3.5)式的第二等式就意味着一步预报误差的条件方差为常数. 这是一个很苛刻的假定. Engle 和 An H Z 等(见文献[45]、[3])都曾指出这一条件的不合理性. 因为在许多经济现象中,对其观测数据所建立的 LAR 模型的残差的条件方差,也即一步预报误差的条件方差,往往呈现出随时间而变化. 换句话说,$E\{e_t^2 | \mathscr{F}_{t-1}^x\}$不是常数. 为便于统计分析,常假定$E\{e_t^2 | \mathscr{F}_{t-1}^x\}$是 t 时刻以前的预报误差 $e_{t-s}^2 (s>0)$的函数. 随机条件方差模型的引进,就是为了刻画这种现象. 下面就我国工业总产值、工业销售收入和社会零售商品物价指数三个经济指标使用1983年1月至1994年9月的月平均数据,进行条件异方差性分析. 先根据经济理论,对数据作一些预处理,包括将原始数据转换为增长率数据和季节调整等. 对预处理后的数据拟合 LAR 模型,对拟合残差序列用 § 6.5的 LM 检验法,对其条件异方差性作检验. 若存在条件异方差性,则建立相应的 ARCH 模型.

(一) 工业总产值:先将工业总产值转化为增长率,再对增长率的数据作季节调整,调整后的数据记为 GI10SAt,用 AIC 准则对该数据拟合 LAR 模型,得到如下 LAR(1)模型:

$$\mathrm{GI10SA}_t = 1.7375 - 0.3657\mathrm{GI10SA}_{t-1} + e_t. \qquad (9.3.6)$$

令

$$\hat{e}_t = \mathrm{GI10SA}_t - 1.7375 + 0.3657\mathrm{GI10SA}_{t-1}$$
$$(t = 2, 3, \cdots, 141).$$

检验$\{e_t\}$是否具有条件异方差性等价于检验如下假设:

$$H_0: E\{e_t^2 | \mathscr{F}_{t-1}^x\} = \sigma^2(\text{常数}).$$

为了使用 LM 检验法,考虑 ARCH 模型

$$\begin{cases} e_t = \varepsilon_t h_t^{1/2}; \\ h_t = \varphi_0 + \sum_{j=1}^{q} \varphi_j e_{t-j}^2, \end{cases} \qquad (9.3.7)$$

其中$\{\varepsilon_t\}$为白噪声序列,满足(3.3.10)式,

$$\varphi_0 > 0,\ \varphi_j \geqslant 0\ (j = 1, \cdots, q),\ \varphi_1 + \cdots + \varphi_q < 1,$$

检验如下假设:

$$H_0: \varphi_1 = \varphi_2 = \cdots = \varphi_q = 0,$$

对立假设为

$$H_1: \sum_{j=1}^{q} \varphi_j \neq 0.$$

利用 LM 检验方法在置信水平 $\beta=0.05$下,依工业总产值数据分析,接受 H_1.

(二) 工业销售收入:作类似于(一)的预处理后,拟合 LAR 模型,得到

$$\text{GI20SA}_t = 2.60 - 0.4138\text{GI20SA}_{t-1} + 0.0366\text{GI20SA}_{t-2} + e_t, \tag{9.3.8}$$

其中"GI20SA_t"表示经预处理后的工业销售收入数据.

令

$$\begin{aligned}\hat{e}_t = \text{GI20SA}_t &- 2.60 + 0.4138\text{GI20SA}_{t-1}\\ &- 0.0366\text{GI20SA}_{t-2} \quad (t = 3,4,\cdots,141).\end{aligned}$$

对残差序列$\{\hat{e}_t\}$,经 LM 检验,在置信水平 $\beta=0.05$下,拒绝 H_0,即认为残差序列$\{\hat{e}_t\}$存在明显的条件异方差性. 于是用 § 7.1(三)所介绍的极大似然估计迭代算法,求得如下 ARCH(2)模型:

$$h_t = 30.88 + 0.3432\hat{e}_{t-1}^2 + 0.0907\hat{e}_{t-2}^2, \tag{9.3.9}$$

由上述三步,即拟合 LAR 模型(9.3.8),对残差序列进行条件异方差性检验,拟合 ARCH 模型(9.3.9),最后得到工业销售收入的 LAR-ARCH 模型

$$\begin{cases}\text{GI20SA}_t = 2.60 - 0.4138\text{GI20SA}_{t-1}\\ \qquad\qquad + 0.0366\text{GI20SA}_{t-2} + e_t;\\ e_t = \varepsilon_t h_t^{1/2};\\ h_t = 30.88 + 0.3432e_{t-1}^2 + 0.0907e_{t-2}^2.\end{cases} \tag{9.3.10}$$

(三) 社会零售商品物价指数:以 P_t 表示社会零售商品物价指数. 由有关经济学理论知,社会零售商品物价指数与上一年同期相比的货币流通量(记为 DTR)有密切关系. 数据分析也表明它们之间存在高度线性相关. 因此,对社会零售商品物价指数数据,我们不是拟合自回归模型,而是拟合以 DTR_t、DTR_{t-1}和 P_{t-1}作为解释变量的混合回归模型,其拟合的模型为

$$\begin{aligned}P_t = 0.017 &+ 3.71 \times 10^{-5}\text{DTR}_t - 3.04 \times 10^{-5}\text{DTR}_{t-1}\\ &+ 0.98\text{P}_{t-1} + e_t.\end{aligned} \tag{9.3.11}$$

对拟合残差$\{\hat{e}_t\}$,经 LM 检验,在置信水平 $\beta=0.05$下拒绝 H_0. 再用极大似然估计迭代算法得到如下 ARCH 模型:

$$h_t = 7.77 \times 10^{-5} + 0.2635\hat{e}_{t-1}^2 + 0.098\hat{e}_{t-2}^2. \tag{9.3.12}$$

这里需要说明的是:估计 ARCH 模型(9.3.12)参数的极大似然迭代算法,是按 Engle(见文献[45])给出的具有 ARCH 误差的回归模型的算法.这一算法的导出与§7.1(三)中介绍的 LAR-ARCH 模型的 MLE 的迭代算法完全类似.

图9.3.1绘出了 h_t 随时间变化的图形.图9.3.2是 h_t 的95%的置信区间.

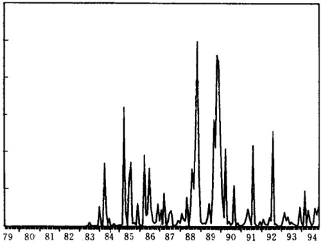


图 9.3.1 h_t 随时间变化图

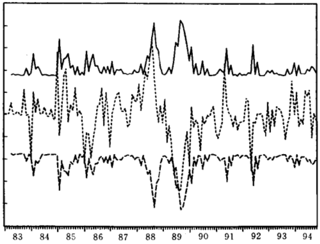


图 9.3.2 h_t 的95%的置信区间

参 考 文 献

[1] An H Z. Non-negative autoregressive models,J Time Series Anal. 1992,13:283～195

[2] An H Z and Chen Z G. On identification of ARMA processes, Appl Prob Special issue. 1986,75～87

[3] An H Z, Chen Z G, and Hannan E J. Autocorrelation, autoregression and autoregressive approximation,Ann Statist. 1982,10:926～936

[4] An H Z, Chen M, and Huang F C. The geometric ergodicity and existence of moments for a class of nonlinear autoregressive models, Statist & Prob Letters ,1997,31:213～224

[5] An H Z and Cheng B. A Kolmogorov-Smirnov type statistic with application to test for non-linearity in time series, Inter Statist Rew. 1991,59:287～307

[6] An H Z and Huang F C. The geometric ergodicity of nonlinear autoregressive models, Statist Sinica ,1996,6:943～956

[7] An Hongzhi, Zhu Lixing and Li Runze. A mixed type test for linearity in time series, 1996, to appear.

[8] Akaike H. A new look at statistical model identification, IEEE Trans. 1974. AC-19:716～723

[9] Ashley R A, Patherson D M and Hinch M J. A diagnostic test for nonlinearity of serial dependence time series fitting errors, J Time Series Anal. 1986,7:165～178

[10] Auestad B and Tjϕstheim D. Identification of nonlinear time series: First order Characterization and order estimation,Biometrika,1990,77:669～687.

[11] Berliner M L. Statisticas, Probability and chaos, Statist Sci,1992,7:69～122

[12] Bierns H J. Uniform consistency of kernal estimation of a regression function under generalized conditions, J Am Staist Assoc, 1983,78:699～707

[13] Bierns H J. Kernal estimators of regression function, Cambridge Univ Press, Advances in Econometrics, 1987

[14] Bollerslev T. Generalized autoregressive conditional heteroskadesticity, J Econometrics, 1986,31:307～327

[15] Bowman A W. An alternative method of cross-validtion for the smoothing ofdensity estimates, Biometrika, 1984,71:353～360

[16] Box G E P and Jenkins G M. Time Series Analysis: Forecasting and Contral, Holden-Day, 1970

[17] Breiman L and Friedman J H. Estimation optimal transformation for multiple regression and correlations (with discussion), J. Am Statist. Assoc, 1985,80:580～619

[18] Breusch T S and Pagan A R. The lagrange multiplier test and its applications to model specification in econometrics, Review of Economic Studies, 1980,XLVII:239～253

[19] Brillinger D R. Time Series: Data Analysis and Theory, Holt,Rinchart and Winston,1975

[20] Brockwell P J and Davis R A. Time Series Analysis: Theoriy and Methods, Springer-ver-

lag, 1987

[21] Broomthad D S and Lowe D. Multivariate functional interpolation and adaptive networks, Complex systems, 2(19):321～355

[22] Burg J P. Maximum entropy spectral analysis, 37th Annual International S. E. G. Meeting, Oklahoma City, Oklahoma.

[23] Buse A. The likelihood ratio, Wold and Lagrange multiplier test: An expository note, Am Statist., 1982, 36:153～157

[24] Chen M and An H Z. A Kolmogorov-Smirnov test of linearity for a class of time series models, Chinese Science bulletin, 1996, 12:1～6

[25] Chen M and An H Z. The Stationarity and hihger-order moments of GARCH model, to appear in Statistica Sinica

[26] Chen M and An H Z. A Kolmogorov-Smirnov test of conditional heteroskadesticity for time series models, to appear in Statist & Prob Letters

[27] Chan K S. Consistency and limiting distribution of the least squares estiomor of a threshold atuoregressive model, Ann Statist, 1995, 21:520～532

[28] Chan K S and Tong H. On the use of deterministic Leyapunov function for the ergodicity of stochastic difference equations, Adv Appl Prob, 1985, 17:666～678

[29] Chan K S and Tong H. On test for nonlinearity in time series analysis, J. Forrcasting, 1986, 5:217～228

[30] Chan K S and Tong H. A note on noisy choas, J Roy Statist Soc, 1994, B56:301～311

[31] Chatterjee S and Yilmag M R. Chaos, Fractals and Statistics, Statist Sci, 1992, 7:49～68

[32] Chen R and Tsay R S. Functional-coefficient autoregressive model, J Am Statist Assoc, 1994, 88:298～308

[33] Chen R and Tsay R S. Nonlinear additive ARX models, J. Am Statist Assoc, 1994, 88: 995～967

[34] Cheng B and Titherington D M. Neural networks: A review from a statistical perspective, Statist Sci, 1994, 9:2～54

[35] Cheng B and Tong H. On consistent non-parametric order determination and choas (with discussion), J Roy Statist. Soc, 1992, B54:427～474

[36] Ciarlet P G. Introduction á l'analyse numérrique matricielle et á l'optimisation, Masson, Paris, 1982

[37] Clutler C D. A review of the theory and estimation of fractal dimension, 1-107, in Dimension estimation and models, ed. by Tong, H, 1994

[38] Collomb G and Hardle W. Strong uniform convergence rate in robust nonparametric time series analysis and prediction: Kernal regression estimation from dependent observations, Stoch Proce Appl, 1986, 23:77～89

[39] Cox D D and Koh E. A smoothing slpine based test of model adequacy in polynomial regression, Ann. Inst. Statist. Math, 1989, 41:383～400

[40] Crowder M J. Maximum likelihood estimation for dependent observations, J. Roy Statist

Soc, 1976, B38: 45～53

[41] Devaney R. An Introduction to Chaotic dynamical systems, 2nd ed, Addison-Wesley, 1989

[42] Diebolt J and Guegan D. Probabilistic properties of the general nonlinear autoregressive process of order one, Technical report, No. 125, L S T A, University Paris VI, 1992

[43] Doob J L. Stochastic Process, John Wiley & Sons, 1953

[44] Elton C and Nicholson M. The ten-year cycle in numbers of the lynx in Canada, J Anim Ecol, 1942, 11: 215～244

[45] Engle R F. Autoregressive conditional heteroskadesticity with estimates of variance of U. K. inflation, Econometrica, 1982, 52: 987～1008

[46] Engle R F, Granger C W S, Rice J and Weiss A. A semiparametroc estimates of the relation between weather and electricity sales, J. Am Statist Assoc, 1986, 81: 310～320

[47] Eubank R L and Speiegelman C H. Testing the goodness of fit of a linear model via nonparametric regression tschniques, J Am Statist Assoc, 1990, 85: 387～392

[48] Feigin P D and Tweedie R L. Random coefficient autoregressive processes — a Markov chain analysis of stationarity and finiteness of moments, J. Time Series Anal, 1985, 6: 1～14

[49] Feigin P D. Maximum likelihood estimation for continuous-time stochastic processes, Adv Appl Prob, 1976, 8: 712～736

[50] Fridman J H and Stuetzle W. Projection pursuit regression, J Am Statist Assoc, 1981, 76: 817～823

[51] Gabr M M and Subba Rao T. The estimation and Prediction of subset bilinear time series models with applications, J Time Series Anal, 1981, 2: 153～171

[52] Gao M Q, Kong N, Xiang J T and Zhang G Q. The prediction of geomagnetic activity maimum and peak time in 22nd solar cycle, The 6th Scientific Assembly, 1989

[53] Grahn T. A conditional least squares approach to bilinear time series estimation, J. Time Series Anal, 1995, 16: 509～529

[54] Granger G W J and Anderson A P. An Introduction to Bilinear Time Series Models, Gottingen, 1978

[55] Guegan D and Diebolt J. Probabilstic properties of the β-ARCH model, Statist Sinica, 1994, 4: 71～88

[56] Hall P and Heyde C C. Martingale Limit Theory and its Application, Academic Press, 1980

[57] Hannan E J. Multiple Time Series, John Wiley & Sons, 1971

[58] Hannan E J and Deistler M. The Statistical Theory of Linear Systems, John Wiley & Sons, 1988

[59] Hannan E J and Quinn B G. The determination of the order of an autoregression, J. Roy Statist Soc, 1979, 41B: 190～195

[60] Hardle W and Hall P. On backfitting algorithm for additive regression models, Statist Neerlandica, 1993, 47: 43～57

[61] Hardle W and Tsybakov A B. Locally polynomial estimstors volatility function. Thechnical reprot, Institut fur statistik und Okonmetric, Humboldt-Universitat zu Berlin, 1994

[62] Hardle W and Vieu P. Kernal regression smoothing of time series, J. Time Series Anal, 1992, 13: 209~232

[63] Hastie T and Tibshirani. Generalized Additive Models. Chapman and Hall. London, 1990

[64] Heckman N E. Spline smoothing in a partly linear model, J. Roy Statist Soc, 1986, B48: 244~248

[65] Hinich M. Testing for Gaussianity and linearity of a atationary time series, J. Time Series Anal, 1982, 3: 169~176

[66] Hjellivik V and Tjφstheim D. Nonparametric test of linearity for time series, Biometrika, 1994

[67] Horn R A and Johnson C. Matrix Analysis, Combridge Univ, Press, 1985

[68] Karlin S and Taylor H M. A First Course in Stochastic Process, Sed Ed Academic Precess, New York, 1975

[69] Keenan P J. A Turkey nonadditivity-type test for time series, Biometrika, 1985, 72: 39~44

[70] Kim W K, Billard L and Basawa I V. Estimation for first-order diagonal bilinear time series model, J Time Series Anal, 1990, 11: 215~229

[71] Klimko L A and Nelson P I. On conditional least squares estimation for stochastic process, Ann Statist, 1978, 629~642

[72] Lagarias J G. The $3x+1$ problem and its generalizations, Am Math Monthly, 1985, 92: 2~23

[73] Li K C. Sliced inverse regression for dimension reduction (with discussion), J Am Statist Assoc, 1991, 86: 316~342

[74] Li T Y and Yorke J A. Period three implies chaos, Am Math Monthly, 1975, 82: 985~992

[75] Luukkonen R, Saikkonen P and Teravirta T. Testing linearity agaigst smooth transition atuoregression, Biometrika, 1988, 75: 491~199

[76] Maravall A. An application of nonlinear time series forecasting, J. Bus Econ Statist, 1983, 1: 66~74

[77] Marsy E and Tjφ sheim D. Non-parametric estimation and identification of ARCH and ARX nonlinear time series: convergence properties and rates.

[78] McCllagh P and Nelder frs J A. Generalized Linear Models, Chapman & Hall, 1989

[79] Mcleod A I and Li W K. Diagnostic Checking ARMA time series models using squared-residual autocorrelations, J time Series Anal, 1983, 4: 269~273

[80] Moran P A P. The statistical analysis of the Canada lynx cycle, Aust J. Zool, 1953, 1: 163~173, 291~298

[81] Muller B and Reinhardt J. Neural Networks: An Introduction, Springer-Verlag, 1990

[82] Nicholls D F and Quinn B G. Random Coefficient Autoregressive Models: An Introduction, Springer-Verlag, 1982

[83] Nummelin E. General Irreduible Markov Chains and Non-negative Operators, Cambirdge

Univ Press,1984

[84] Ozaki T. Nonlinear time series models and aynamical systems, Handbook of Statistics, Vol 5 (ed. Hannan E J, Krishnaish P R and Rao M M),North-holland,1985

[85] Ozaki T and Oda H. Nonlinear time series model identification by Akaike's information criterion,Proc IFAC workshop of Information and System, Compiegn,Frence, 1978

[86] Park B U and Marron J S. Comparision of data-driven bandwith selecters, J. Am Statist Assoc, 1990,85:66~72

[87] Peitgen H U and Richter R H. The Beauty of Fractals, Springer-Verlag, 1986

[88] Rademo M. Empirical choice of histograms and kernal density estimators, Scand J. of Statist,1982,9:65~78

[89] Robinson P M . Non-parametric estimation for time series models, J. Time Series Anal, 1983,4:185~208

[90] Rosenblatt M. A central limil theorem and strong mixing condition, Proc Nat Acad Sci, 1956,42:43~47

[91] Saikkonen P and Luukkonen R. Lagrange multiplier test for testing nonlinearity in time series models, Scand J Statist, 1988,15:55~68

[92] Sarkovskii A N. On a theorem of G B Birkhoff, Dopovidi Akad Nauk Ukvain RSR Ser A, 1967,429~432

[93] Schick A. On asymptotically efficient estimation in semiparametric models, Ann Statist, 1986,14:1139~1151

[94] Sesay S A O and Subba Rao T. Difference equations for higher-order momnets and cummulants for the bilinear time series model BL (p,0,p,1),J. Time Series Anal ,1991,12:159~176

[95] Shibata R. Selection of the order of an autoregressive model by Akaike's information criterion,Biometrika, 1976,63:117~126

[96] Singh R S and Ullah A. Nonparametric time series estimation of joint DGP, conditional DGP and vector autoregression,Econometric Theory,1986,1

[97] Stone C J. Consistent nonparametric regression, Ann Statist,1977,5:595~645

[98] Subba Rao T and Gabr M M . A test for linearity of stationary time series, J. Time Series Anal,1980,1:145~158

[99] Subba Rao T and Gabr M M. An Introduction to Bispectral Analysis and Bilinear Time Series Models, Lecture in Statistics, No 24, Springer-Verlag, 1984

[100] Teravirta T, Saikkonen P and Luukkonen R. Power of the neural network linearity test, J time Series Anal,1993,14:209~220

[101] Tjφstheim D. Some doubly stochastic time series models, J. Time Series Anal, 1986,7:52~72

[102] Tjφstheim D. Non-linear time series and Markov chains, Adv Appl Prob, 1990,22:587~611

[103] Tong H. Some comments on the Canadian lynx data-with discussion, J. Roy Stst Soc,

1977,a140:432～435,448～468

[104] Tong H. A view on nonlinear time series model building,In Time Series (Ed Anderson O O), North Holland,1980

[105] Tong H. Threshold Models in Nonlinear Time Series Analysis, No,20,Springer-Verlag, 1983

[106] Tong H. Non-linear Time Series, Oxford Univ,Press, 1990

[107] Tong H. A personal overview of nonlinear time series analysis from a chaos perspective, An invited lecture at the 15th nordic conference on mathematical statistics, Lund, Sweden, August,1994

[108] Tong H and Lim K S. Threshold autoregression, limit cycles and cyclical data (with discussion), J Roy Stat Soc,1980,B42:245～292

[109] Tran L T. Nonparametric function estimation for time series by local average estimators, Ann Statist,1993,21:1040～1057

[110] Truong Y K . A nonparametric framework for time series analysis, New Directions in Time Series, Springer-Verlag,1993

[111] Truong Y K and Stone C J. Nonparametric function estimation involving time series, Ann Statist,1992,20:77～97

[112] Truong Y and Stone C J. Semiparamtric time series regression, J. Time Series Anal, 1994,15:405～428

[113] Tsay R S. Nonlinearity tests for time series, Biomerika, 1986,73:237～252

[114] Tweedie R L. Sufficient conditions for ergodicity and geometric ergodicity of Markov chain on a general state space, Stoch Proc appl. 1975,3:385～403

[115] Tweedie R L. Criteria for classifying general Markov chains, Adv Appl Prob, 1983,20: 191～196

[116] Volterra V. Theory of Functionals, Blackie, 1930

[117] Walters P. An Introduction to Ergodic Theory,Springer-Verlag, 1982

[118] Zhu Lixing and An Hongzhi. A testing approach for nonlinearity in regression models, J. of Math. 1992,4:398～397(in Chinese)

[119] 马志明,安万福. 独立和分布为连续的充要条件. 应用数学学报,1985,8:204～209

[120] 文志英,井竹君,分形几何和分维数简介. 数学实践与认识,1995,No. 4:20～29

[121] 史久恩,施永年,卫纪范. 动力统计长期数值天气预报的进展. 北京:科学出版社,1989

[122] 孙传忠,安鸿志,吴国富. ARCH 模型及其应用与发展. 数理统计与应用概率,1995,10:63～68

[123] 卢祖帝. 关于双重时序 AR-MA 模型存在平稳解的充要条件. 应用数学学报,1994,17:374～387

[124] 卢祖帝. 关于双重时序 AR-MA 模型的相关结构及与 ARMA 模型的比较. 系统科学与数学,1996,16

[125] 安鸿志,陈兆国,杜金观,潘一民. 时间序列的分析与应用. 北京:科学出版社,1983

[126] 陈敏,安鸿志. ARCH (p) 模型的严平稳遍历性和高阶矩. 科学通报,1995,40:2118～2123

[127] 范文宁,邓集贤. 门限自回归模型参数估计的渐近性质. 应用概率统计,1986,2:259～266
[128] 项静恬等. 动态和静态数据处理. 北京:气象出版社,1991
[129] 张尧庭,方开泰. 多元统计分析引论. 北京:科学出版社,1983
[130] 张桂清. 用自激励门限自回归分析方法模拟和预报太阳黑子数的长期变化. 天体物理学报,1988,8
[131] 陶宗英,何声武,汪嘉冈译. 希尔伯特空间中的平稳序列. 上海:上海科学技术出版社,1963
[132] 常学将,刘维奇. AR 模型识别及其参数的高阶 Yule-Walker 估计. 应用数学学报,1989,12:218～227
[133] 常学将,陈敏,王明生. 时间序列分析. 北京:高等教育出版社,1993
[134] 盛昭瀚,王涛,刘德林. 非线性时间序列模型的稳定性分析. 北京:科学出版社,1993
[135] 谢衷洁. 时间序列分析. 北京:北京大学出版社,1988

名 词 索 引

一～二画

三画

四画

五画

六画

七画

八画

九画

十画

十一画

十二画以上